Classification of Synthetic Aperture Radar Images

Michael Hagedorn

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

11.08.2004

ABSTRACT

In this thesis the maximum *a posteriori* (MAP) approach to synthetic aperture radar (SAR) image analysis is reviewed. The MAP model consists of two probability density functions (PDFs): the likelihood function and the prior model. Contributions related to both models are made.

As the first contribution a new likelihood function describing the multilook three-polarisation intensity SAR speckle process, which is equivalent to the averaged squared amplitude samples from a three-dimensional complex zero-mean circular Gaussian density, has been derived. This PDF is a correlated three-dimensional chi-square density in the form of an infinite series of modified Bessel functions with seven independent parameters. Details concerning the PDF such as the estimation of the PDF parameters from sample data and the moments of the PDF are described. The new likelihood function is tested against simulated and measured SAR data.

The second contribution is a novel parameter estimation method for discrete Gibbs random field (GRF) prior models. Given a quantity of sample data, the parameters of the GRF model, which comprise the values of the potential functions of individual cliques, are estimated. The method uses an error function describing the difference between the local model PDF and the equivalent estimated from sample data. The concept of "equivalencies" is introduced to simplify the process. The new parameter estimation method is validated and compared to Besag's parameter estimation method (coding method) using GRF realisations and other sample data. In most cases the new parameter estimation method performs better than Besag's method.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Phil Bones for his help and guidance in this research. His interest and enthusiasm over the years has been great motivation. Further thanks go to Dr David Pairman and Dr Stephen McNeill of Landcare Research who provided valuable knowledge in image processing and synthetic aperture radar imaging. Many people gave me expert advice on the research which was much appreciated, so many thanks to: Professor Rick Millane, Professor Peter Gough, Professor Mike Steel, Dr Mike Hayes, Dr Heather North, Stella Belliss, Dr Qing Wu. I would especially like to acknowledge Peter Smith's invaluable contribution to Chapter 7. He helped me tackle difficult integrals and always with a laugh.

Then I want to thank all my friends who shared the research, coffee, waves etc. so a few names (in alphabetical order): Bart and Marleen, Christoph and Barbara, Hangi, Hannes and Bettina, Hayden and Tanya, Kai, Nadir, Pieter, Timo and Sonja, Wili and Sarah, the crew from Café 101...

The following scholarships and awards were gratefully received: Doctoral Scholarship from Landcare Research, Lincoln, New Zealand; Community Trust Doctoral Scholarship from University of Canterbury, Christchurch, New Zealand; Royal Society of New Zealand, Canterbury Branch, Travel Award.

Thanks to my parents and family and extended family for their encouragement and support. Especially to Lara and Sven who arrived while this thesis was in progress and gave me many excuses to pause and gather words. Finally thanks to Britta for being so nice.

CONTENTS

Part I

Introduction and Background

Chapter 1

INTRODUCTION

An image is a two dimensional lattice system made up of pixels each having some value (such as for example the gray level value). The data represented in the image of a scene is in most cases easily understood by a human expert even though the images are often corrupted by some noise. The human brain can identify objects of interest and their location within an image without too much difficulty. Programming a computer to perform the same task has proven very difficult and only very specialised image analysis tasks have been successfully implemented. Having a computer do the image analysis has great advantages such as working at constant quality with high speed for a long time, not having to train human experts, archiving reproducible results, etc. With constantly increasing computing power available the implementation of image analysis on computers is becoming more and more practical. A basic task in image analysis is image classification. One type of image classification is assigning a label value to each pixel of the image data. The label value represents a property of the imaged scene which is of interest. Examples of label values could be: forest, water, bare ground and grass areas; human, animal and machine; hot and cold. A classification method produces a label image (image of the same size as the data image with label values at each pixel) indicating the content and the arrangement of that content within the image.

In this thesis the Bayesian approach to image analysis is used. This is a stochastic model approach based on the maximum *a posteriori* (MAP) estimate. The *a posteriori* probability density function (PDF) is a conditional density giving the likelihood of a particular label image given a noisy data image. Using the Bayesian equations this *a posteriori* PDF can be split in two parts, the likelihood function and the prior model. The likelihood function describes the relationship between the label value and the noisy data in the form of a conditional PDF, it returns the likelihood of the noisy image data given the label image. In the simplest form a likelihood function could be such that it returns higher values if the gray level value is high and the label value stands for "hot" than if the gray level value is high and the label value stands for "cold" (for example in an infrared image). This type of likelihood function operates on a single pixel gray level value while a more complex likelihood function may analyse the texture of an area with regard to a label value. The likelihood function is therefore very specific, first with regard to the image data supplied and secondly with regard to the desired label values.

The prior model in the *a posteriori* PDF is a way of incorporating prior knowledge of the underlying label image into the classification. The prior density will return higher values for a label image that is known to be a likely classification outcome. For example if the classification is to find square areas of different content in the data image, a correct prior model would return higher values when the label image is made up of square areas than otherwise. One such model is the Markov random field (MRF), or a special case of it the Gibbs random field (GRF) PDF. These models are based on the spatial Markov property and associate a higher probability to label images made up of homogeneous areas. The size and shape of the homogenous areas can be modified to some degree via the specific GRF model type and model parameters. It is important to note that the prior model in this case only operates on the label image and knows nothing about the noisy image data.

The MAP approach to SAR image classification was chosen in preference to other methods for two main reasons. The first reason is that the stochastic model in the form of a likelihood function is based on the physical interaction of the radar signal with the ground cover. The SAR speckle noise arising from this interaction is well understood and can therefore be implemented into the model. The second reason is that the MAP approach allows prior knowledge about the resulting classification to be implemented. This can be used to guide the typically difficult classification of SAR images.

The two main contributions of this thesis are in the area of the likelihood function and prior model. The first contribution is the development of a new likelihood function for multilook three-polarisation intensity synthetic aperture radar (SAR) data [Hagedorn et al. 2004]. The second contribution is a new parameter estimation method for the model parameters of the Gibbs random field (GRF) prior model.

The newly developed likelihood function is specific to multilook three-polarisation intensity SAR data. The speckle noise of this type of SAR data has not been modelled before. The density is a three dimensional correlated chi-square density and in mathematical terms models averaged intensity samples from a three dimensional complex zero-mean circular Gaussian distribution and therefore applications in other fields may exist. As a motivation for this work stands primarily the deeper understanding of the polarimetric speckle process. The model can be used to separate the speckle noise from any other process affecting the signal such as texture. A SAR imaging satellite planned for launch in early 2004 (RADARSAT-2) will, in one imaging mode, provide this type of data (averaged intensity values of the three polarisations rather than the full complex covariance matrix) [Staples and Hornsby 2002].

The new parameter estimation method developed in this thesis is based on minimising an error function with regard to the GRF model parameters. The error function measures the difference between GRF model characteristics and the same characteristics from a specific label image. The motivation for developing this new parameter estimation method was to get a deeper understanding of the GRF model and the characteristics of the data it describes. Another reason was that alternate parameter estimation methods failed to give satisfactory results for some specific label images.

Using accurate GRF model parameters for the prior model in a MAP classification will yield better results giving rise for the need of a good parameter estimation method.

The thesis is structured in four parts. Part I (Chapters 1 to 6) gives the introduction and the background material required in this thesis. Part II (Chapters 7 and 8) and Part III (Chapters 9 and 10) contain the new contributions given in this thesis comprising theory (Part II) and application (Part III). In Part IV (Chapter 11) conclusions are given and the results are discussed. All abbreviations used throughout the thesis are listed in the index.

Part I is organised as follows. In Chapter 2 the theory of SAR imaging systems and radar signals is provided. After a general introduction to SAR, real aperture strip-map radar systems are described leading to a description of SAR itself. The basic features of radar signals such as frequency and polarisation are introduced. Then the ground interaction of the radar signal is explained.

Chapter 3 gives the background of the probability theory that is required in later chapters. The MAP estimate is explained in some more detail here.

In Chapter 4 stochastic models for different types of SAR signals, for example multilook intensity or polarimetric complex SAR signals, are derived. The origin of speckle noise is explained and the most basic method of reducing the effect of speckle noise, multilook processing, is described. A short note is given on the simulation of speckle noise.

Chapter 5 gives an overview of the temporal and spatial Markov property. This leads to the introduction of random fields and the Gibbs random field (GRF). The conditions ensuring that the GRF is a MRF are given. For this the neighbourhood system and cliques are defined. Then the potential function is introduced which leads to the energy function and thus to the MRF. Finally details of the potential function are discussed and issues with regard to sampling from a GRF and parameter estimation are raised.

In Part II Chapter 7 the new likelihood function is derived. This likelihood function provides a joint density for three intensities measured by an n-look monostatic polarimetric SAR system, which is equivalent to the average of squares of n samples from a correlated, three-dimensional, complex, zero-mean circular Gaussian distribution. Special cases are given to theoretically validate this new joint density. A method of parameter estimation from intensity samples is laid out and the moments of the joint density are derived based on the characteristic function.

Chapter 8 provides the theory of the new parameter estimation method for GRF model parameters. The method presented is based upon the minimisation of a cost function with regard to the model parameters. The cost function gives a value for the difference between the local model PDF and a PDF estimated from the data set. Both PDFs are discrete and have a set number of states. It is shown that the model does not distinguish between some local configurations and these can be combined into a single state which is called here an 'equivalency'. The equivalencies reduce the complexity of the model by a large amount and make the comparison between model and data practical.

In Part III Chapter 9 the new likelihood function, which is derived in Chapter 7, is analysed and

validated. First a proof for convergence of the infinite sum term in the joint density is given. Then the new likelihood function is compared to simulated data. Afterwards the method of estimating the parameters in the joint density is discussed and examples are given. Then real SAR data is compared to the new PDF and finally discussions and conclusions concerning this chapter are presented.

In Chapter 10 the novel method of parameter estimation for the GRF models, which is introduced in Chapter 8, is tested and compared to Besag's parameter estimation method. GRF realisations with known parameter values are used to estimate the parameters. Different model configurations are used in this process. For comparison Besag's parameter estimation method is also implemented. Then three test label images are given. Two are hand drawn two class (binary) label images and the third one is a five class label image generated from geographic information systems (GIS) data. The GRF model parameters are estimated and compared to estimates obtained with Besag's parameter estimation method. The quality of the estimated parameter sets is assessed by way of a MAP classification, using the estimated parameters in the GRF prior model, of noisy data versions of the test label images. The result from the classification is measured against the true test label images. The results of this are discussed and conclusions given.

In the last part, Part IV, discussion and conclusions concerning the thesis are given.

Chapter 2

SYNTHETIC APERTURE RADAR

In this chapter some background information about synthetic aperture radar (SAR) is presented. The chapter is organised as follows: After a general introduction to SAR, attention is directed towards strip-map systems. A description of real aperture strip-map radar leads to a description of SAR itself. The basics features, frequency and polarisation of radar signals are discussed. Then the interaction of the radar signal with the ground surface is explained.

SAR is a remote sensing technique which is separate from other methods of remote sensing. It is a coherent microwave system providing its own illumination and therefore can operate at day and night and see through cloud cover. Depending on the wavelength of the system, it penetrates foliage or even the ground surface to some depth. This gives SAR a great advantage over the classical optical remote sensing tools. Multiple wavelength and polarisation SAR imaging systems are being used. Some of the uses of SAR are biomass estimation, sea ice analysis and oil spill detection [Barni et al. 1995; Derrode et al. 2001; Espedal 1999; Hara et al. 1995; Lee et al. 1994b; Saatchi and Moghaddam 1999; Thompson 2001]. Interferometric SAR is gaining in importance for generating elevation models.

The processing of the collected data from a SAR is split into preprocessing and postprocessing. Preprocessing is often performed on board the SAR platform (satellite or airplane) and consists of signal conditioning, range and azimuth compression and the most basic of noise reduction multi-look processing and auto-focussing. Postprocessing, performed on the resulting SAR image, is made up of several operations such as projections and resampling of the data in order to obtain a ground referenced SAR image scaled and oriented with regard to some global geographic map grid.

SAR was discovered in the 1950s. It developed from strip-map, real aperture radar which was used to image the ground by repeatedly emitting a radar beam perpendicular from a moving platform and measuring the backscattered signal. The along-track resolution of such a system is limited by the aperture size; much better resolution was achieved with the use of SAR. The underlying idea behind SAR is to compress many pulse returns into one pixel.

Different forms of SAR systems exist, providing various types of resolutions and image properties. One distinguishes between bistatic and monostatic (also called reciprocal) SAR. In a bistatic SAR the transmitting and receiving antennas are separate while in a monostatic SAR they are the same (or fixed to the same platform). Most commercial SAR systems are monostatic due to the

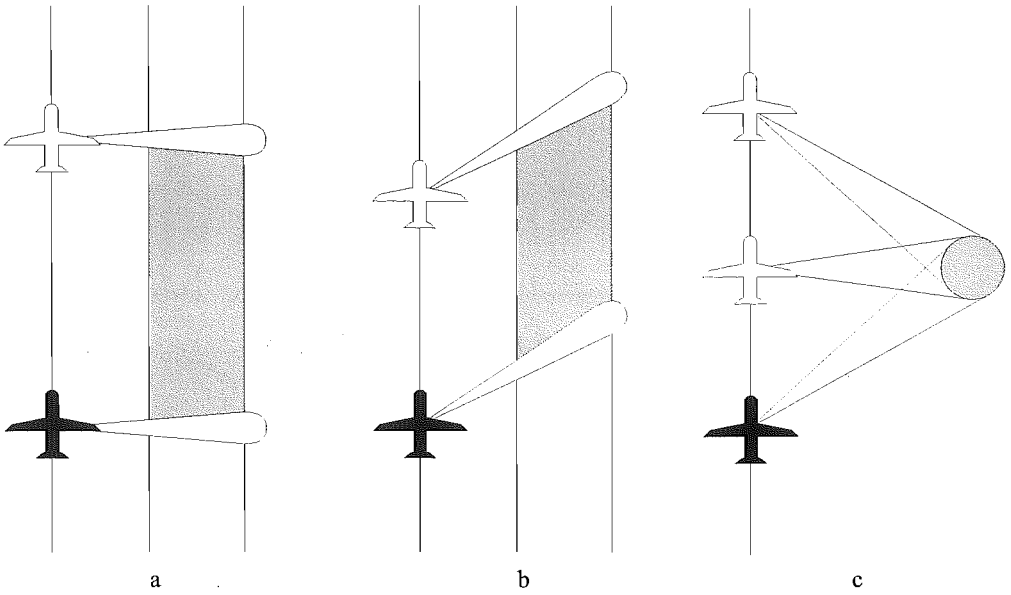


Figure 2.1 Synthetic aperture modes: (a) broadside strip-map, (b) squinted strip-map and (c) spotlight.

reduced complexity and cost. A variety of fixed and steered look angles (illumination direction of the radar beam with respect to the flight path) have been investigated. The most commonly used however are the fixed look angle, in most applications perpendicular to the flight path (broadside strip-map SAR), or the steered squint angle, illuminating the same ground target from different directions (spotlight SAR, see Figure 2.1).

The system discussed in this thesis (and the most commonly used in commercial applications) is broadside strip-map SAR. However most of the presented methods derived here can be applied to any coherent imaging data, such as that arising in laser imaging and sonar imaging (without polarisation issues in the latter case).

The contributions of this thesis are mainly concerned with the statistics of the final SAR image, after range and azimuth compression. Therefore only a basic overview is given to understand how these statistical properties arise (for detailed discussions about SAR, the reader is referred to Curlander and McDonough [1991] and Ulaby [1982b]).

To understand the functioning of such a SAR system it is useful to discuss the real aperture radar imaging system first.

2.1 STRIP-MAP REAL APERTURE RADAR IMAGING

Figure 2.2 shows the general geometry of a side-looking radar. A platform (typically an airplane or satellite) with the radar is flown along a straight path. The radar is directed perpendicular to the flight path and below the horizon by a fixed angle called the look angle. The direction of flight is called along-track or azimuth direction while the perpendicular direction to this on the ground is called the range direction. The radar typically emits chirped electromagnetic (EM) signal pulses of bandwidth B at set intervals. The EM pulse interacts with the ground depending on the

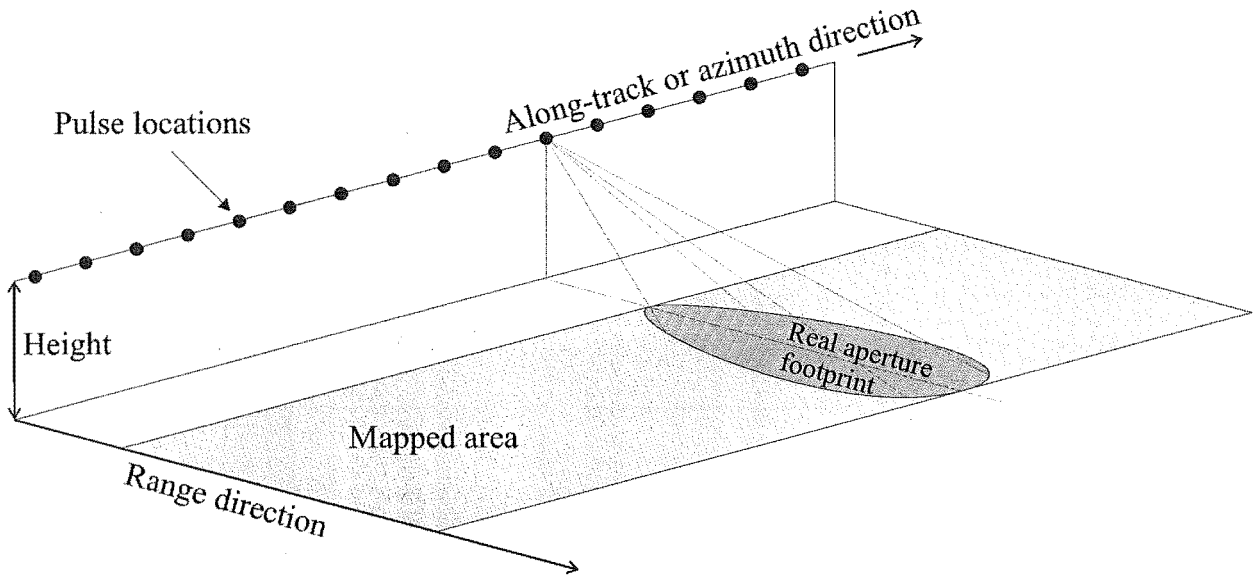


Figure 2.2 The layout of a side-looking radar system.

radar and ground properties. The important radar properties are typically the radar frequency (or equivalently wavelength), antenna polarisation, range bandwidth and look angle. The important ground properties affecting the echo signal are the dielectric constant, roughness and local slope of the ground. After the interaction some of the radar signal is reflected back to the antenna, compressed in range and the delay time and envelope echo strength are measured. The delay time between the pulse emission and the echo reception is dependent on the distance of the target from the platform. Each detected echo therefore produces a one-dimensional scalar function of range, a "line" in the image. As the platform moves along the track, lines are combined to form a real aperture image.

The range resolution of a real aperture radar, defined as the 3dB width of the range compressed pulse, is

$$\delta_{\text{range}} = \alpha_w \frac{c}{2B}, \tag{2.1}$$

where c is the propagation speed of the signal and is well approximated by the speed of light and B is the bandwidth of the signal. The term α_w is a constant and compensates for the effect of any weighting or windowing introduced to reduce range sidelobes of the radar signal (typical values for α_w are around 1). Therefore a larger bandwidth signal enables a better range resolution.

The azimuthal resolution of the real aperture is dependent on the beam width which depends on the signal frequency f and the antenna (aperture) length D and is given by

$$\delta^{\text{real}}_{\text{azimuth}} = \alpha_w \frac{c x_{\text{range}}}{fD}. \tag{2.2}$$

This indicates that for a high azimuthal resolution, a long antenna and a high frequency are needed. In order to obtain an azimuthal resolution of 1km at L-band ($\lambda = 20$cm) with a typical satellite setup, a corresponding aperture length of $D = 160$m is required. For the same satellite with an

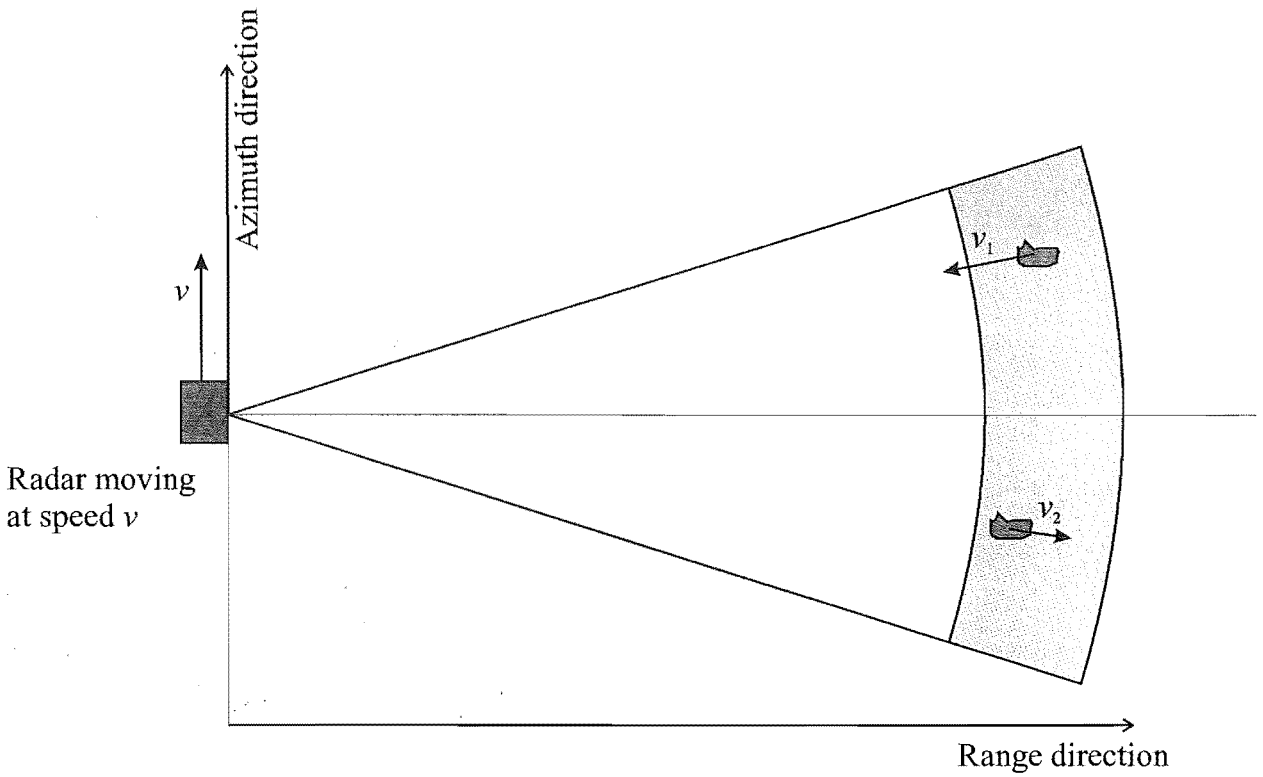


Figure 2.3 Two targets are illuminated by a radar travelling at speed v. Two arrows at the targets indicate the relative speed components v_1 and v_2 of the objects experienced by the radar signal.

aperture of the length of $D = 10.7$m, an azimuthal resolution of $D/2 = 5.4$m can be obtained using SAR [Curlander and McDonough 1991]. For more details on real aperture radar the reader is referred to Curlander and McDonough [1991] and Hawkins [1996].

2.2 STRIP-MAP SYNTHETIC APERTURE RADAR IMAGING

The key observation that led to the development of SAR was that two point targets at the same range (separated by a small angle) have different relative speeds with regard to the moving radar platform. This difference in radial velocity results in two different Doppler shifts in a reflected EM non-chirped pulse. This effect enables the detection of the location of a target within a range-resolution cell and the system to distinguish between multiple targets within a range-resolution cell. Figure 2.3 shows the top view of two targets illuminated by the radar travelling at speed v. One of the targets is ahead of the radar platform, the other one behind. Two arrows at the targets indicate the relative speed components v_1 and v_2 of the objects experienced by the radar signal. These will induce two Doppler shifts in the return signal.

A strip-map SAR system operates in essentially the same way as the strip-map real aperture radar. The difference is that the data is collected, stored and processed in a coherent manner. In order to obtain a high azimuth resolution a large antenna is synthesised. This is done by coherently integrating (summing) in an appropriate manner the received echoes from the imaged target field.

The azimuthal resolution can then be approximated by

$$\delta_{\text{azimuth}}^{\text{synthetic}} = \frac{D}{2} \tag{2.3}$$

where D is the aperture length. This result is counterintuitive since a smaller aperture will result in a higher resolution. Some assumptions have been made in deriving Equation 2.3, but most resolutions achieved by SAR systems approach this limit. The resulting image has an azimuthal resolution that is independent of the range and wavelength. The processing involved is essentially a two-dimensional, range-variant, matched filter operation. Different methods have been developed to implement the filter but a discussion of its design goes beyond the scope of this thesis. For a detailed description of synthetic aperture processing, see Curlander and McDonough [1991].

After all the synthetic aperture processing, a data set is obtained which is mapped onto a regular grid, for which every site (i.e. SAR image pixel) corresponds to a rectangular ground resolution cell. The cell has width and length corresponding to the range and azimuthal resolution. The data at every site is, for example, a single scalar value (like intensity) or a polarimetrically sensitive complex matrix (Stokes matrix), indicating the radar reflectivity of the ground resolution cell. Typically resolutions range from below one meter to many meters in modern SAR systems. It is this data set which is the subject of study throughout the reminder of the thesis.

2.3 RADAR SIGNALS

In this section the properties of radar signals are discussed. An EM signal is characterised by its frequency, phase, amplitude and polarisation. From Maxwell's equations, the expression for the complex electric field vector $\mathbf{E}$ at a given location z in the propagation direction and at time t is given by

$$\mathbf{E}(z,t) = \mathbf{A}\exp\left(j(k'z - \omega t + \phi)\right), \tag{2.4}$$

where $\mathbf{A}$ is the amplitude vector (which includes any polarisation), k' is the wavenumber in the medium and the angular frequency is given by ω. The phase shift referenced to some datum is ϕ. Furthermore,

$$\omega = 2\pi f_c = 2\pi c/\lambda \tag{2.5}$$

$$k' = \sqrt{\varepsilon_r}k = 2\pi\sqrt{\varepsilon_r}/\lambda, \tag{2.6}$$

where f_c denotes the carrier frequency and λ the wavelength. The relative permittivity is given by ε_r and the vacuum wavenumber by k.

2.3.1 Signal frequency and bandwidth

The transmitted pulse is normally a linear FM pulse (a chirp) with center frequency f (for SAR in the order of 1-15GHz). The bandwidth of the chirp is typically small compared to the center

Band name	Frequency range (GHz)	Wavelength (cm)
P	0.22 - 0.39	136 - 77
UHF	0.3 - 1	100 - 30
L	1 - 2	30 - 15
S	2 - 4	15 - 7.5
C	4 - 8	7.5 - 3.75
X	8 - 12.5	3.75 - 2.40
Ku	12.5 - 18	2.40 - 1.67
K	18 - 26.5	1.67 - 1.18
Ka	26.5 - 40	1.18 - 0.75

Table 2.1 Commonly used frequency bands and their wavelengths.

frequency (around 10MHz) and determines the range resolution (see Equation 2.1). Table 2.1 shows the most commonly used frequency bands, band names and wavelengths of SAR systems.

2.3.2 Signal polarisation

Many modern SAR systems send out polarised EM pulses and measure the polarisation of the echo. The polarisation of an EM signal, as given in Equation 2.4, refers to the direction of the amplitude vector **A** at a given time. In SAR four main polarisation modes are used: horizontal, vertical, elliptical and circular. Horizontally polarised means that the direction of the amplitude vector is fixed in the horizontal plane and correspondingly for the vertical polarisation mode. Elliptically polarised means that the vector **A** is rotating with time around the z direction following the outline of an ellipse. The circularly polarised mode is a special case of this, where length of **A** stays constant. For a detailed treatment of signal polarisation in radar see Ulaby and Elachi [1990] and Mott [1992].

2.4 SURFACE INTERACTION WITH THE RADAR SIGNAL

This section gives a brief overview of the surface interaction of a radar signal. The interaction of the radar pulse with the ground cover is generally called scattering and depends on the signal properties such as amplitude, polarisation and frequency, and the ground cover properties such as dielectric constant, roughness and local slope. Roughness is defined with regard to the wavelength of the radar signal. The same surface can appear rough to a radar signal with a short wavelength and smooth to a signal with a longer wavelength. Note that in a monostatic SAR a rougher surface has a higher mean reflectivity than a relatively smooth surface (the signal is not reflected back to the antenna).

Different scattering modes such as surface scattering, volume scattering and Bragg scattering exist. In reality, a mixture of scattering modes is encountered [Lewis et al. 1998; Ogilvy 1991; Raney 1998; Ulaby 1982a, b].

2.4.1 Speckle

Speckle, which is also called fading, is an inherent property of the coherent nature of SAR imaging, rather than a system noise feature. The speckle noise in SAR images gives them a pronounced granular appearance and makes the analysis of SAR data difficult. Mathematical models for speckle noise have been developed and are discussed in Chapter 4. A wide variety of algorithms to remove the effect of speckle noise have been developed the most commonly used is multi-look processing as described in Section 4.1.2. Other speckle filtering methods are listed in Frost et al. [1982]; Lee [1983]; Lee et al. [1991, 1994c]; Lopès and Séry [1997]; Lopes and Touzi [1988]. More detailed information about speckle can be found in Goodman [1975, 1976, 1986].

2.5 SHADOW, FORESHADOW AND LAYOVER IN SAR

Besides the inherent speckle noise of SAR images, other features distort the appearance of the image. Some are a direct result of the SAR system viewing ground with topographic features such as mountains and canyons at an angle other than normal (i.e. the look angle is $< 90°$). The distortions are called: shadow, foreshadow and layover. The look angle varies with range starting with the steepest angle at the closest range to the most oblique angle at the farthest range. Therefore the effects are range-dependent. Figure 2.4 shows these effects. A projection of representative corresponding radar response intensity values are given below the graph in ground cells. Note that black indicates zero intensity and occurs in the shadow. Features on the mountain slope facing the radar are projected into ground cells closer to the radar, this feature is called foreshadow. The figure shows the extreme case of foreshadow, i.e. the top of the mountain is in the same ground cell as some part of the foot of the mountain, a phenomenon known as layover.

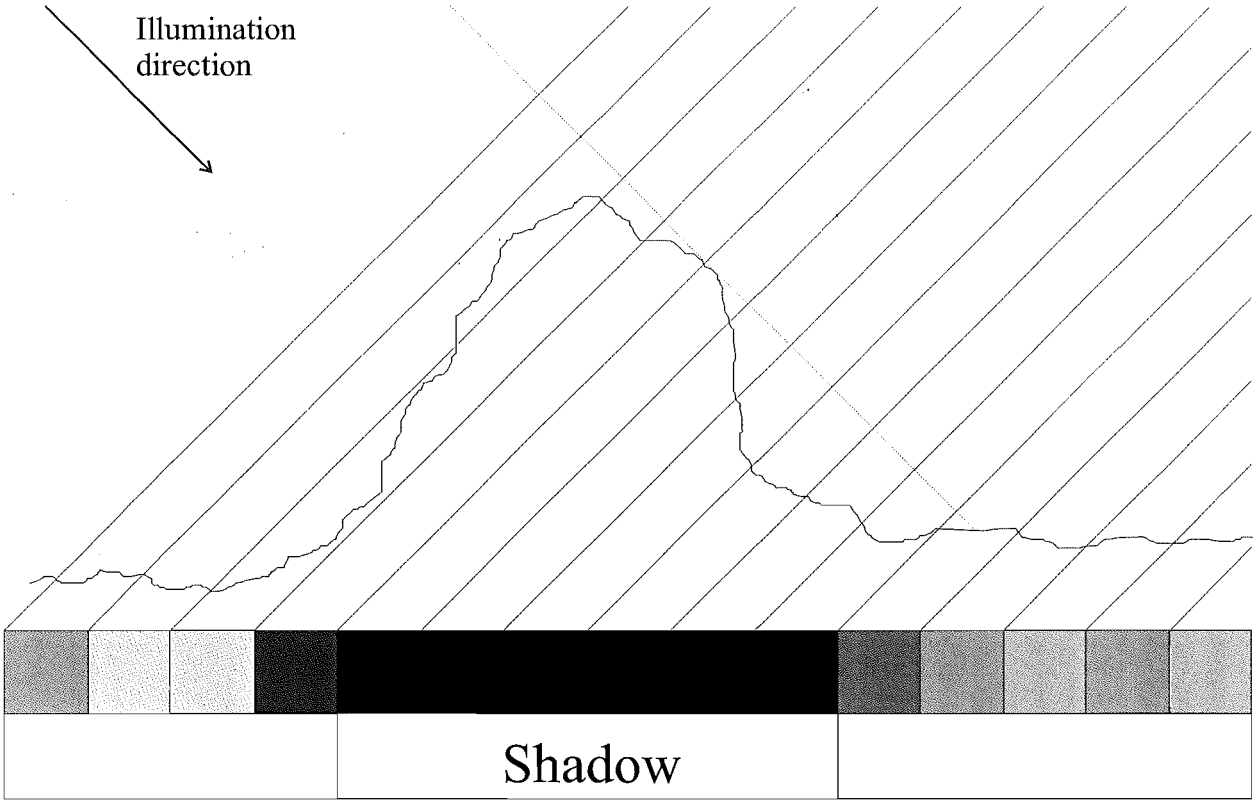


Figure 2.4 An example showing shadow, foreshadow and layover.

Chapter 3

PROBABILITY AND ESTIMATION

In this chapter probability and estimation theory are reviewed to provide background for the remainder of the thesis. For a detailed and complete definition of probability theory the reader is referred to Papoulis [1991] and Stark and Woods [1994] and for estimation theory to Kay [1993].

3.1 PROBABILITY THEORY

Probability theory can be defined (from set theory) by three axioms. Papoulis [1991] defines them as follows:

Let $S = \{s_1, \ldots, s_k\}$ be the sample space, the set of all possible outcomes s_i of an experiment. An event A occurs during a trial if it contains the element s_i. Assign to each event A, a number $\mathrm{P}(A)$ which is called the probability of the event A. This number is chosen to satisfy the following three conditions:

1. The probability of an event A is between zero and unity: $0 \leq \mathrm{P}(A) \leq 1$.
2. The probability of the sample space is unity: $\mathrm{P}(S) = 1$.
3. If the events A and B are two mutually exclusive events, then $\mathrm{P}(A + B) = \mathrm{P}(A) + \mathrm{P}(B)$.

3.1.1 Conditional probability

The conditional probability of A given B is the probability of the event A occurring given that event B occurred. This conditional probability $\mathrm{P}(A|B)$ is defined by:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A, B)}{\mathrm{P}(B)} \tag{3.1}$$

where $\mathrm{P}(B)$ is assumed to not be 0 and $\mathrm{P}(A, B)$ is the probability of events A and B both occurring, the joint probability. Further, by applying the Bayes's theorem

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)}. \tag{3.2}$$

If the events A and B are statistically independent then $\mathrm{P}(A, B) = \mathrm{P}(A)\mathrm{P}(B)$ and $\mathrm{P}(A|B) = \mathrm{P}(A)$.

3.1.2 Distribution and density function

In order to define the cumulative distribution and the density function, the concept of a random variable must be introduced. A random variable X assigns a number to every outcome s: $X(s)$. $X(s)$ has to follow the two conditions [Papoulis 1991]:

1. The set $\{X \leq x\}$ is an event for every x.

2. The probability of the events $\{X = \infty\}$ and $\{X = -\infty\}$ is equal to zero, i.e. $\mathrm{P}(X = \infty) = 0$ and $\mathrm{P}(X = -\infty) = 0$.

The cumulative distribution function is defined as

$$\mathrm{F}(x) = \mathrm{P}(X \leq x). \tag{3.3}$$

The derivative of the cumulative distribution function is called the probability density function (PDF) (or density function)

$$\mathrm{f}(x) = \frac{d\mathrm{F}(x)}{dx}. \tag{3.4}$$

3.1.2.1 Properties of probability density functions

The stochastic modelling in this thesis is mainly based on PDFs, thus it is useful to reflect on some of the properties of a PDF.

An inherent property of a PDF is that its integral over the whole range is equal to one:

$$\int_{-\infty}^{\infty} \mathrm{f}(x)dx = 1. \tag{3.5}$$

The probability $\mathrm{P}(x_1 < X \leq x_2)$ is given by

$$\mathrm{P}(x_1 < X \leq x_2) = \int_{x_1}^{x_2} \mathrm{f}(x)dx. \tag{3.6}$$

With $x_1 = x$ and $x_2 = x + \Delta x$ it follows that

$$\mathrm{P}(x \leq X \leq x + \Delta x) \simeq \mathrm{f}(x)\Delta x \tag{3.7}$$

given that Δx is sufficiently small. Thus $\mathrm{f}(x)$ is proportional to the probability of X within a small interval of x.

Bayes's theorem for densities is

$$\mathrm{f}(x|y) = \frac{\mathrm{f}(y,x)}{\mathrm{f}(y)} = \frac{\mathrm{f}(y|x)\mathrm{f}(x)}{\mathrm{f}(y)} \tag{3.8}$$

where $\mathrm{f}(y, x)$ is the joint density defined as

$$\mathrm{f}(x, y) = \frac{\partial^2 \mathrm{F}(x, y)}{\partial x \partial y} \tag{3.9}$$

with

$$\mathrm{F}(x, y) = \mathrm{P}(X \leq x, Y \leq y). \tag{3.10}$$

The denominator of Equation 3.8 can be written as

$$\mathrm{f}(y) = \int_{-\infty}^{\infty} \mathrm{f}(x, y) dx \tag{3.11}$$

which is also known as the total probability. If the PDF's are discrete Equation 3.11 becomes

$$\mathrm{f}(y) = \sum_{\forall x} \mathrm{f}(x, y). \tag{3.12}$$

If the random variables X and Y are independent, then $\mathrm{f}(x, y) = \mathrm{f}(x)\mathrm{f}(y)$ and $\mathrm{f}(x|y) = \mathrm{f}(x)$.

3.1.3 Statistical quantities

In this section the most important statistical quantities are defined.

3.1.3.1 Expected value

The mean or expected value of the random variable X is

$$\mu_x = \mathrm{E}(X) = \int_{-\infty}^{\infty} x \mathrm{f}(x) dx \tag{3.13}$$

3.1.3.2 Variance

The variance is defined as

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 \mathrm{f}(x) dx \tag{3.14}$$

where σ_x is called the standard deviation of X. The following identity can be obtained

$$\sigma_x^2 = \mathrm{E}\left((X - \mu_x)^2\right) = \mathrm{E}\left(X^2 - 2X\mu_x + \mu_x^2\right) = \mathrm{E}(X^2) - 2\mu_x \mathrm{E}(X) + \mu_x^2, \tag{3.15}$$

hence

$$\sigma_x^2 = \mathrm{E}\left(X^2\right) - \mathrm{E}^2\left(X\right). \tag{3.16}$$

3.1.3.3 Moments

The kth moment of X is defined as

$$m_k = \mathrm{E}\left(X^k\right) = \int_{-\infty}^{\infty} x^k \mathrm{f}(x) dx. \tag{3.17}$$

The kth central moment of X is defined as

$$\eta_k = \mathrm{E}\left((X-\mu)^k\right) = \int_{-\infty}^{\infty} (x-\mu)^k \mathrm{f}(x) dx. \tag{3.18}$$

Note that $\mu_x = \eta_1$, $\sigma_x^2 = \eta_2$ for X.

3.1.3.4 Covariance and correlation

The covariance between the two random variables X and Y is defined as

$$C_{xy} = \mathrm{E}((X-\mu_x)(Y-\mu_y)) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y). \tag{3.19}$$

The correlation coefficient is the normalised covariance with regard to the standard deviations of the two variables, thus

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}. \tag{3.20}$$

Further, $|C_{xy}| \leq \sigma_x \sigma_y$ and therefore $|\rho_{xy}| \leq 1$ [Papoulis 1991].

3.2 THE MAXIMUM *A POSTERIORI* ESTIMATE

A standard task in image processing is: given a noisy image $\mathbf{s}$, calculate a processed version $\mathbf{r}$. The processing takes a variety of forms, for example noise filtering, classification, or segmentation. For example: given $\mathbf{s}$, which is corrupted by speckle noise, calculate $\mathbf{r}$ where each pixel in $\mathbf{r}$ has a label indicating a class of ground cover of the corresponding pixel in $\mathbf{s}$. Classes of ground cover could be of the type: "bare ground", "trees", "water", "built up area", and so on. This process is called classification.

The statistical approach to this problem is: calculate an estimate $\hat{\mathbf{r}}$ maximising the probability of $\mathbf{r}$ given $\mathbf{s}$, thus

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} [\mathrm{f}(\mathbf{r}|\mathbf{s})] \tag{3.21}$$

where "arg" denotes the argument of the following function, in this case $\mathbf{r}$. Using Bayes's theorem for PDFs in Equation 3.8 leads to

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} \left[\frac{\mathrm{f}(\mathbf{s}|\mathbf{r})\mathrm{f}(\mathbf{r})}{\mathrm{f}(\mathbf{s})}\right]. \tag{3.22}$$

Since f(**s**) is known and independent of **r**,

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} [\mathrm{f}(\mathbf{s}|\mathbf{r})\mathrm{f}(\mathbf{r})] . \tag{3.23}$$

The estimate $\hat{\mathbf{r}}$ as defined in Equation 3.23 is called the maximum *a posteriori* estimate. The term f(**s**|**r**) is called the likelihood function and f(**r**) the prior model. All noise filtering and classification algorithms used in this thesis are based on Equation 3.23. Several steps are involved in gaining an estimate $\hat{\mathbf{r}}$: deriving an appropriate likelihood function and prior model, estimating model parameters, and optimisation. For a detailed explanation of estimation theory the reader is referred to Kay [1993].

In this thesis the noisy data that is supplied is simply referred to as the data image. The results of processing are referred to as estimate images, or simply estimates, e.g. the MAP estimate, the label estimate, etc.

3.2.1 The likelihood function

The likelihood function, f(**s**|**r**), can be thought of as an inverse (backwards) mapping operation. In other words, given the result **r** how probable is the data **s**. For example: if the label for a given pixel in **r** corresponds to "bare ground", then the likelihood function will return a higher value if the corresponding pixel in **s** contains the radar response of "bare ground" rather than "trees". The likelihood function in this case is dependent on the class characteristics (such as mean and standard deviation for example) and the features of the inverted imaging process (speckle noise statistics).

A range of possible likelihood functions for SAR images are discussed in Chapter 4.

3.2.2 The prior model

The prior model, f(**r**), determines how likely a resulting image **r** is. If prior knowledge about the result is available it can be incorporated by means of the prior model. For example, if it is known that the result is made up of square patches of fields, then it is possible to design a prior model which assigns a higher probability to results that contain square patches. It is important to know that the prior model is only dependent on **r** and has no information about the noisy data **s**.

In this thesis a Markov random field (MRF) model is used as a prior model. This model is used to assign a higher probability to results **r** made up out of homogeneous areas of ground cover. On a local basis in **r** this corresponds to: a site in **r** is likely to have the same value as its surrounding site values. The theory of MRFs is described in Chapter 5.

If no prior knowledge is available the prior density has the form of a uniform distribution (every outcome has the same likelihood) and Equation 3.23 becomes

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} [\mathrm{f}(\mathbf{s}|\mathbf{r})] . \tag{3.24}$$

Under these circumstances $\hat{\mathbf{r}}$ is said to be the maximum likelihood (ML) estimate.

Chapter 4

MODELLING AND SIMULATING SAR DATA

In this chapter stochastic models are derived for different modes of SAR data. Speckle noise is defined and the most basic method of reducing the effect of speckle noise, multilook processing, is described. A short note is given on how to simulate basic speckle noise. Simulating SAR data has an important role in verifying the speckle filter and classification methods. The underlying speckle free-signal is known and by means of simulation quantitative assessments can be obtained. For a detailed discussion on speckle see [Goodman 1975, 1976, 1986; Raney 1998]

4.1 SPECKLE MODEL

Underlying all of the models presented here is speckle noise. It is therefore important to understand how this type of noise arises and what its effects are. The theory of speckle noise has been thoroughly researched and the models are well understood. It is generally accepted that the speckle noise present in the SAR data recorded from most natural scenes is described by a circular Gaussian distribution [Lee et al. 1994b]. Thus the complex radar return signal from a ground resolution cell within a natural scene has a circular Gaussian distribution. This property arises from the random distribution of scatters in the ground resolution cell (see Figure 4.1) and the central limit theorem (sometimes described by means of a random walk).

Goodman [1976] describes the origin of speckle thus: "When a radar illuminates a surface that is rough on the scale of radar wavelength, the return signal consists of waves reflected from many randomly-distributed elementary scatterers (or facets) within a resolution cell. The distance between the scatterers and the receiver vary, therefore the received waves, although coherent in frequency, are no longer coherent in phase. A strong signal is received if waves add relatively constructively; a weak signal, if the waves are out of phase."

This speckle model fails when there are not enough scatterers in the ground resolution cell or a single scatterer dominates the response. This is often the case in urban areas and when other man-made structures are present. These contain many large smooth right-angle configurations such as house walls with regard to the ground. These features, due to their geometry, act as dihedral corner reflectors. Most of the signal is reflected straight back to the antenna and they appear as bright intensity values in the SAR image.

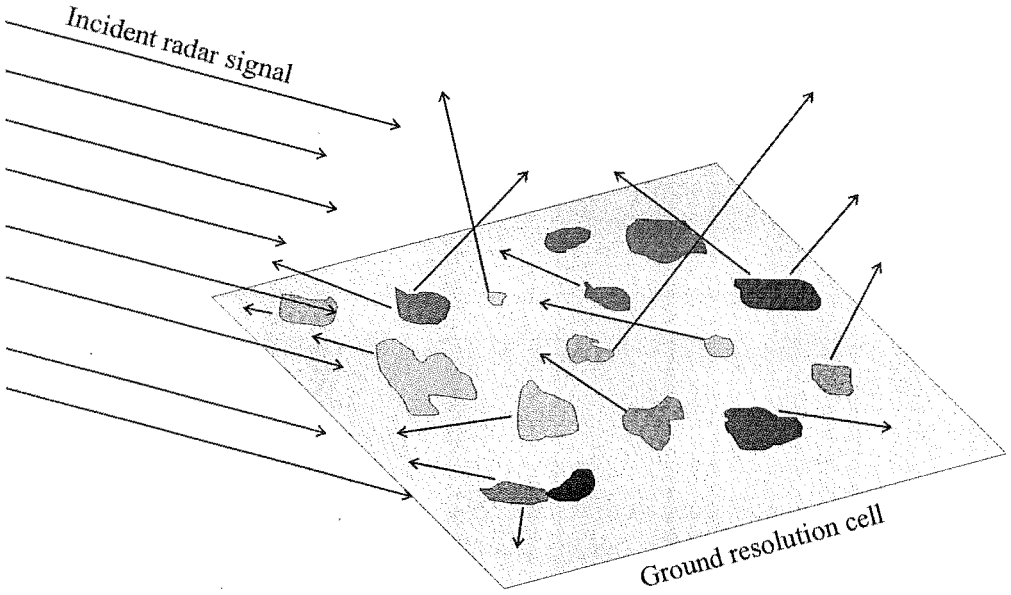


Figure 4.1 Schematic of random scatterers of a radar signal in a ground resolution cell.

Frequently an estimate of some property of the ground cover is to be gained from the SAR data collected from a scene. If for example intensity values of the radar return signal of a homogeneous ground area are measured, this area appears as an area with a range of intensity values in the image domain (due to the presence of speckle noise). Each ground resolution cell corresponds to a pixel with an intensity value in the image domain. Since the ground cover in this example is assumed to be homogeneous, the intensity values can be seen as samples from a single PDF. A good estimate might therefore be obtained by calculating the mean of the intensity values from a group of adjacent pixels, given there are a sufficient number of them. In practice however the problem lies in determining whether samples are from the same ground cover type or a different ground cover type. Mixtures of ground cover types have to be considered as well. In the following sections the known PDFs for different SAR data modes are introduced.

4.1.1 Single-look complex data model (the circular Gaussian model)

If the number of scatterers within a resolution cell is large, the real and imaginary parts of the radar signal are zero-mean Gaussian distributed with the same standard deviation (but uncorrelated); this is by argument of the central limit theorem. Often the minimum number of necessary scatterers is given as 5, but Raney [1998] suggests that 10 are required to assure convergence of the statistics. The phase is therefore totally random or uniformly distributed. One way of deriving these properties is by means of random walk theory as follows (see Goodman [1976]).

4.1.1.1 The random walk to speckle

Let $\tilde{X}$ be the complex return signal from a ground resolution cell. $\tilde{X}$ is the sum of individual complex signals a_k from many elementary scatterers within the ground resolution cell. Let $|a_k|$ and

ϕ_k be the amplitude and phase of the kth elementary scatterer respectively. It follows that

$$\tilde{X} = \sum_k |a_k| \exp(j\phi_k). \tag{4.1}$$

If the scatterers are distributed randomly in the ground resolution cell and the surfaces are rough with regard to the wavelength of the signal then the following assumptions hold:

1. The amplitude $|a_k|$ and the phase ϕ_k of the kth contributor are statistically independent of each other and of all other $|a_l|$ and ϕ_l with $l \neq k$.

2. The phases ϕ_k are uniformly distributed over $(0, 2\pi)$, $\forall k$.

From Equation 4.1 it may appear that the greater the number of elementary scatters within a resolution cell, the greater the signal strength. However, the returned signal strength is clearly limited by the finite amount of transmitted radiation reaching the cell. Each elementary scatterer is associated with an effective scattering cross-section with respect to the incoming radiation and direction of detection of the re-radiated radiation. This cross-section is approximately inversely proportional to the number of elementary scatters in the resolution cell and therefore the returned signal given by Equation 4.1 is bounded.

Given these assumptions, results from the classical random walk theory can be used [Goodman 1975, 1976]. Figure 4.2 shows an example of a random walk. The elementary signal returns are shown as vectors in the complex domain.

Let $\tilde{X} = X_\mathrm{R} + jX_\mathrm{I}$ and $\mathbf{W} = [X_\mathrm{R}\ X_\mathrm{I}]^T$ be the vector of the real and imaginary parts of this signal. Then the density of $\mathbf{W}$ is, f($\mathbf{W}$), a zero-mean circular Gaussian, thus

$$\mathrm{f}(\mathbf{W}) = \left[(2\pi)^2 |\mathbf{C}|\right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{W}^T\mathbf{C}^{-1}\mathbf{W}\right) \tag{4.2}$$

where $\mathbf{C}$ is the covariance matrix of $\mathbf{W}$ defined as $\mathbf{C} = \mathrm{E}[\mathbf{W}\,\mathbf{W}^T]$. The real and imaginary parts of $\tilde{X}$ have the same variance and are uncorrelated (this is a property of a circular complex Gaussian and direct result from the two assumptions given above) thus

$$\mathbf{C} = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_X^2 \end{bmatrix}. \tag{4.3}$$

$|\mathbf{C}|$ is the determinant of $\mathbf{C}$. The phase of $\tilde{X}$ has a uniform density and the magnitude and the intensity have Rayleigh and negative exponential density respectively [Goodman 1975].

This model describes statistically pure speckle or what is sometimes called fully developed speckle. It is encountered in many natural scenes and gives a good approximation in most remaining cases. One exception is where a single scatterer dominates the response, like in the case of a dihedral corner reflector. Another complication is that speckle can be spatially correlated due to the imaging system. Correlated models are not discussed further in this thesis (see Raney [1998] for

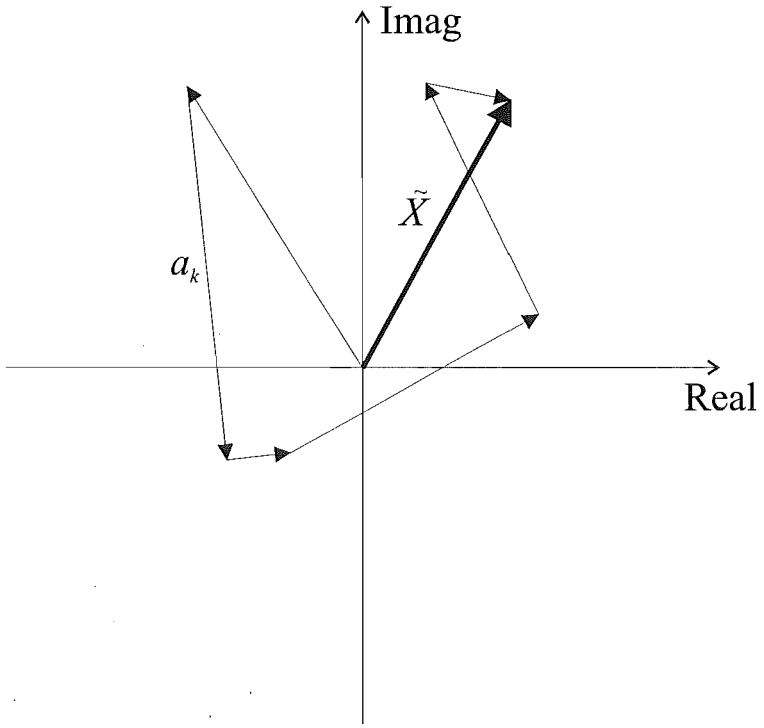


Figure 4.2 Random walk example in the complex domain where a_k is the kth elementary scatterer and $\tilde{X}$ the resulting measured return signal.

details on correlated speckle). First order speckle models, such as PDF's on individual site values in conjunction with simple texture models for textured ground covers, have been shown to yield good results in many SAR image analysis cases [Baraldi and Parmiggiani 1995; Cook et al. 1996; Fjørtoft et al. 1996; Liu et al. 1996; Nezry et al. 1991].

Speckle statistics are sometimes affected by textural features of the scene imaged. This means that a first order texture model is often used in conjunction with the speckle noise model. This topic is discussed in more detail in Chapter 6.

4.1.2 Multi-look processing

The simplest form of speckle reduction is multi-look processing and most commercial data is provided in a multi-look form. Multi-look processing is essentially averaging n independently sampled images of the same scene. The process reduces the variance of the data at the expense of resolution. The independently sampled images are in general obtained from the same SAR data set by exploiting the holographic property of SAR data (see next section). The resolution of an n-look image is reduced by the factor n, while the variance is one nth of that of the corresponding single look image. The resulting image appears less noisy.

Multi-look processing is often applied such that the resolution in one dimension only (azimuth or range) is reduced and thereby a SAR image results with a constant range and azimuth resolution. In this case every pixel in the SAR image corresponds to a square ground resolution cell.

A non-integer number of looks can be generated by using overlapping sections of the data in Fourier space.

4.1.2.1 Holographic properties of SAR data

Consider the two-dimensional object reflectivity function

$$\xi(x,y) = |\xi(x,y)| \exp[j\ \phi(x,y)] \tag{4.4}$$

where $\phi(x,y)$ is the phase of $\xi(x,y)$ and x and y are the spatial coordinates [Hawkins 1996]. The Fourier transform of Equation 4.4 is given by

$$\mathcal{F}\{\xi(x,y)\} = \mathcal{F}\{|\xi(x,y)|\} * \mathcal{F}\{\exp[j\ \phi(x,y)]\} \tag{4.5}$$

where $\mathcal{F}\{\cdot\}$ and $*$ indicate the two-dimensional Fourier transform and two-dimensional convolution operator respectively. The phase of natural scenes is known to be highly random (uniformly distributed) and therefore $\mathcal{F}\{\exp[j\ \phi(x,y)]\}$ has a broad spectrum. Thus, the random phase modulates the magnitude information over a wide region of Fourier space. This enables the recovery of an image estimate from any section of $\mathcal{F}\{\xi(x,y)\}$. The location of the section does not affect the target surface and shape information but the size of the section determines the reduction in resolution. The same underlying principles are used to generate holograms [Munson and Sanz 1984].

4.1.3 Intensity and amplitude model for multi-look data

In multi-look processing the final image intensity is formed by averaging n independent intensities (see Section 4.1.2). Let the intensity I_{X} and the amplitude A_{X} of the n-look data be

$$I_{\mathrm{X}} = A_{\mathrm{X}}^2 = \sum_{i}^{n} \left(X_{\mathrm{R}i}^2 + X_{\mathrm{I}i}^2\right)/n. \tag{4.6}$$

The intensity of a four-look image for example is the sum of four intensity samples from the same zero-mean complex circular Gaussian density given in Equation 4.2. The sum of squared Gaussian samples has a chi-square density (Abramowitz and Stegun [1972] page 940), thus the intensity for n-look data has a chi-square density of the form

$$\mathrm{f}(I_{\mathrm{X}}) = \frac{n^n I_{\mathrm{X}}^{n-1}}{\Gamma(n)\sigma^{2n}2^n} \exp\left(-\frac{nI_{\mathrm{X}}}{2\sigma^2}\right) \tag{4.7}$$

where $\Gamma(\cdot)$ is the Gamma function. The mean and the variance of I_{X} are $\mu_{\mathrm{I_X}} = 2\sigma^2$ and $\sigma_{\mathrm{I_X}}^2 = 4\sigma^4/n$ respectively. The variance can therefore be expressed as a function of the mean $\sigma_{\mathrm{I_X}}^2 = \mu_{\mathrm{I_X}}^2/n$. This confirms that the variance decreases proportionally with the number of looks n and therefore the

effect of speckle is reduced. Note that for $n = 1$ the standard deviation of I_X is the same as the mean.

Besides the number of looks n, which is a property of the processing of the SAR data, there is the effective number of looks (ENL) which can be measured from the data [Lee et al. 1994c]. The ENL value is defined as

$$\text{ENL} = \mu_{I_X}^2 / \sigma_{I_X}^2. \tag{4.8}$$

This value is often used to measure the effect of speckle filtering because as the ENL increases the effect of speckle decreases. Further, if the ENL is less then n this indicates underlying texture or other variation in the scene. This is due to texture increasing the standard deviation.

The amplitude n-look data has a Rayleigh density of the form [Lee et al. 1994c]

$$\text{f}(A_X) = \frac{2n^n A_X^{2n-1}}{\Gamma(n)\sigma^{2n}} \exp\left(-\frac{nA_X^2}{\sigma^2}\right) \tag{4.9}$$

with the mean

$$\mu_{A_X} = \frac{\Gamma(n+1/2)}{\Gamma(n)}\sqrt{\sigma/n} \tag{4.10}$$

and variance

$$\sigma_{A_X}^2 = \left(n - \frac{\Gamma^2(n+1/2)}{\Gamma^2(n)}\right)\frac{\sigma^2}{n}. \tag{4.11}$$

4.1.4 Correlated complex polarised model

Many modern SAR systems are able to separate horizontally (H) and vertically (V) polarised components of the transmitted and received signals. Four combinations are therefore available for a polarimetric SAR system, e.g. HV refers to a horizontally transmitted and vertically received signal. In a monostatic system the HV response is identical to the VH response and hence three polarisations are measured (HH, VV and HV). The SAR signal is therefore three-dimensional.

The measured data of a monostatic polarised SAR systems therefore consists of three complex amplitudes $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$, one for each polarisation HH, VV and HV respectively. Let $\tilde{\mathbf{W}} = [\tilde{X}, \tilde{Y}, \tilde{Z}]^T$ and the corresponding vector of real and imaginary parts be denoted by $\mathbf{W} = [X_R, X_I, Y_R, Y_I, Z_R, Z_I]^T$. Section 4.1.1 describes the properties of a one dimensional single look complex SAR signal. The same assumptions given there, for the one dimensional signal, hold for each of the three complex signals. Therefore the three polarisation complex signal has a three-dimensional zero-mean complex circular Gaussian distribution.

The circular Gaussian property of the data means that the real and imaginary parts of, for

example, the $\tilde{X}$ and $\tilde{Y}$ polarisations are correlated as [Goodman 1963; Lee et al. 1994b]:

$$\begin{aligned}
\mathrm{E}(X_\mathrm{R}X_\mathrm{R}) &= \mathrm{E}(X_\mathrm{I}X_\mathrm{I}) = \sigma_X^2 \\
\mathrm{E}(Y_\mathrm{R}Y_\mathrm{R}) &= \mathrm{E}(Y_\mathrm{I}Y_\mathrm{I}) = \sigma_Y^2 \\
\mathrm{E}(X_\mathrm{R}X_\mathrm{I}) &= \mathrm{E}(X_\mathrm{I}X_\mathrm{R}) = 0 \\
\mathrm{E}(Y_\mathrm{R}Y_\mathrm{I}) &= \mathrm{E}(Y_\mathrm{I}Y_\mathrm{R}) = 0 \\
\mathrm{E}(X_\mathrm{R}Y_\mathrm{R}) &= \mathrm{E}(X_\mathrm{I}Y_\mathrm{I}) = c_{XYA} \\
\mathrm{E}(X_\mathrm{I}Y_\mathrm{R}) &= -\mathrm{E}(X_\mathrm{R}Y_\mathrm{I}) = c_{XYC}.
\end{aligned} \tag{4.12}$$

where the subscript "A" indicates "auto" and "C" indicates "cross" with respect to the real and imaginary parts. Similarly for the combinations of $\tilde{X}$ and $\tilde{Z}$, $\tilde{Y}$ and $\tilde{Z}$. The covariance matrix $\mathbf{C} = \mathrm{E}[\mathbf{W}\,\mathbf{W}^T]$ therefore takes the form

$$\mathbf{C} = \left[\begin{array}{cc|cc|cc}
\sigma_X^2 & 0 & c_{XYA} & -c_{XYC} & c_{XZA} & -c_{XZC} \\
0 & \sigma_X^2 & c_{XYC} & c_{XYA} & c_{XZC} & c_{XZA} \\
\hline
c_{XYA} & c_{XYC} & \sigma_Y^2 & 0 & c_{YZA} & -c_{YZC} \\
-c_{XYC} & c_{XYA} & 0 & \sigma_Y^2 & c_{YZC} & c_{YZA} \\
\hline
c_{XZA} & c_{XZC} & c_{YZA} & c_{YZC} & \sigma_Z^2 & 0 \\
-c_{XZC} & c_{XZA} & -c_{YZC} & c_{YZA} & 0 & \sigma_Z^2
\end{array}\right]. \tag{4.13}$$

The covariance matrix $\tilde{\mathbf{C}}$ for the complex variables $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$ is defined by $\tilde{\mathbf{C}} = \mathrm{E}[\tilde{\mathbf{W}}\,\tilde{\mathbf{W}}^\dagger]$ (where $\cdot^\dagger$ is the Hermitian transpose or conjugate transpose) and takes the form

$$\tilde{\mathbf{C}} = \mathrm{E}\left[\begin{array}{ccc}
\tilde{X}\tilde{X}^* & \tilde{X}\tilde{Y}^* & \tilde{X}\tilde{Y}^* \\
\tilde{Y}\tilde{X}^* & \tilde{Y}\tilde{Y}^* & \tilde{Y}\tilde{Z}^* \\
\tilde{Z}\tilde{X}^* & \tilde{Z}\tilde{Y}^* & \tilde{Z}\tilde{Z}^*
\end{array}\right] \tag{4.14}$$

where $\cdot^*$ is the conjugate operator. $\tilde{\mathbf{C}}$ is a conjugate symmetrical (Hermitian) matrix and positive semi-definite. It therefore has positive eigenvalues. $\tilde{\mathbf{C}}$ can be written with regard to the real variances and covariances defined in Equation 4.13 as

$$\tilde{\mathbf{C}} = 2\left[\begin{array}{ccc}
\sigma_X^2 & c_{XYA} + j\,c_{XYC} & c_{XZA} + j\,c_{XZC} \\
c_{XYA} - j\,c_{XYC} & \sigma_Y^2 & c_{YZA} + j\,c_{YZC} \\
c_{XZA} - j\,c_{XZC} & c_{YZA} - j\,c_{YZC} & \sigma_Z^2
\end{array}\right]. \tag{4.15}$$

The circular zero-mean Gaussian distribution of the signals $\mathbf{W}$ has the form

$$\mathrm{f}(\mathbf{W}) = \left[(2\pi)^6\,|\mathbf{C}|\right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{W}^T\mathbf{C}^{-1}\mathbf{W}\right), \tag{4.16}$$

and the complex version of $\tilde{\mathbf{W}}$ is

$$\mathrm{f}(\tilde{\mathbf{W}}) = \frac{1}{\pi^3 \left|\tilde{\mathbf{C}}\right|} \exp\left(-\tilde{\mathbf{W}}^{\dagger}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{W}}\right). \tag{4.17}$$

For a detailed analysis of these type of multivariate Gaussian distribution see Goodman [1963].

4.1.4.1 The complex Wishart density

The multilook processing of polarimetric data is performed on the covariance matrix of the SAR data (an averaging of the complex data vectors $\tilde{\mathbf{W}}_i$ would approach zero since the samples are from a zero mean distribution). The resulting n-look covariance matrix $\tilde{\mathbf{R}}$ is therefore defined by

$$\tilde{\mathbf{R}} = \frac{1}{n}\sum_{i}^{n} \tilde{\mathbf{W}}_i \tilde{\mathbf{W}}_i^{\dagger}. \tag{4.18}$$

The complex n-look covariance matrix $\tilde{\mathbf{R}}$ is known to have a complex Wishart density [Goodman 1963; Lee et al. 1994b]

$$\mathrm{f}(\tilde{\mathbf{R}}) = \frac{n^{qn}\left|\tilde{\mathbf{R}}\right|^{n-q} \exp\left(-n\mathrm{Tr}\left(\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{R}}\right)\right)}{K(n,q)\left|\tilde{\mathbf{C}}\right|^{n}} \tag{4.19}$$

with

$$K(n,q) = \pi^{(1/2)q(q-1)}\Gamma(n)\cdots\Gamma(n-q+1) \tag{4.20}$$

where q is the dimension of the vector $\tilde{\mathbf{W}}$ (in our case $q = 3$, in the bistatic case $q = 4$) and $\mathrm{Tr}(\cdot)$ denotes the trace of the matrix $(\cdot)$. The complex Wishart density based on the circular Gaussian assumption has been successfully used in SAR image analysis, for examples see [Ferro-Famil et al. 2001; Lee et al. 1994a, 1999, 2001; Lopès and Séry 1997].

4.1.4.2 The two dimensional intensity density

In some cases the SAR data set is not a complete set of all three polarisations or intensity values rather than the complex values are supplied. For this reason Lee et al. [1994b] derived the joint density for intensities from two polarisations from polarimetric SAR. This density written in the notation used in this thesis is

$$\mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}}) = \frac{\exp\left(\dfrac{\sigma_Y^2 I_{\mathrm{X}} + \sigma_X^2 I_{\mathrm{Y}}}{2(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)}\right)(I_{\mathrm{X}}I_{\mathrm{Y}})^{(n-1)/2}}{\left(\sigma_X^2\sigma_Y^2 - c_{XYA}^2 - c_{XYC}^2\right)\Gamma(n)2^{n+1}\left(c_{XYA}^2 + c_{XYC}^2\right)^{(n-1)/2}} \mathrm{I}_{n-1}\left(\frac{\sqrt{c_{XYA}^2 + c_{XYC}^2}\sqrt{I_{\mathrm{X}}I_{\mathrm{Y}}}}{\sigma_X^2\sigma_Y^2 - c_{XYA}^2 - c_{XYC}^2}\right), \tag{4.21}$$

where $I_{n-1}(\cdot)$ is the modified Bessel function of order $n-1$. In Chapter 7 the density for intensities from all three polarisations is derived (a three dimensional PDF on polarimetric intensity SAR data). Further the two dimensional case in Equation 4.21 is shown to be a special case of the three dimensional one and the notation in Lee et al. [1994b] is linked to the one used here.

4.1.5 Multi-frequency model

Some modern SAR systems acquire a multipolarisation and multifrequency data set. A common data mode is three polarisations (HH, VV and HV) and three frequency bands. The data set therefore consist of a nine dimensional complex vector or an eighteen dimensional real vector for each site corresponding to a ground resolution cell.

Different frequency bands highlight different ground cover properties. If the features on the ground are in the order of a wavelength a higher response can be expected than if the features are far smaller than the wavelength of the radar signal. Additional frequencies therefore improve the ability to classify the ground cover.

The simplest and most common model used for multilook, multipolarisation and multifrequency data sets is the multivairate Gaussian distribution. This model assumes that all data subsets are correlated and have a Gaussian shape. This however is only true if the data has a high number of looks. In general the higher the number of looks the closer is the shape of the distribution to a Gaussian distribution.

4.2 SIMULATING SPECKLE

In order to simulate speckle, samples from one of the above speckle models, depending on the type of data desired, have to be created. The easiest way to create samples for any of the above SAR speckle models is to sample from a multivariate zero-mean Gaussian density as described in Equation 4.2 or Equation 4.16 for polarimetric data. Multitlook speckle can be generated by averaging independent samples from these densities and intensity, amplitude or covariance values can be formed as defined in Equation 4.6 or Equation 4.18. Two examples of simulated single look intensity SAR data will be given in Chapter 9.

Chapter 5

THEORY OF MARKOV RANDOM FIELDS

Markov random field (MRF) models have been used in image analysis since the mid eighties. The MRF model in image analysis is mostly used in conjunction with Bayesian statistical methods in the form of a prior model or sometimes as part of the likelihood function (see Chapter 6 for references on this topic). The model reflects the fact that neighbouring pixel (site) values are likely to depend on each other. For example: if all neighbours of a site have the same value, say class A, then the site itself is more likely to have the value of class A rather then a different one. If from the noisy data a site value is as much likely to belong to class A as to class B this rule can guide the decision by using the class values from neighbouring sites.

The MRF model used most often in image analysis is the Gibbs random field (GRF) model. This model is based on the Gibbs distribution. It is very flexible in the way that one can use arbitrary lattice systems as long as a neighbourhood system can be defined. It can therefore be applied to one, two, three or n-dimensional space. In order to ensure that the GRF model defined is a MRF one has to follow some rules. These rules are laid out in the following sections. Besag [1974] proves that under these circumstances the resulting model is indeed a MRF model.

For a detailed description of MRF models in image analysis and optimisation with regard to them see Winkler [1995], Dubes and Jain [1989] and Geman and Geman [1984].

First in this chapter the temporal and spatial Markov properties are discussed, then random fields are defined and the GRF is introduced. The conditions which ensure that the GRF is a MRF are laid out, in the process defining formally the neighbourhood and cliques. Then the potential function is introduced which leads to the energy function and thus to the MRF. An example of a GRF model is given. In the final sections details of the potential function are given and issues with regard to sampling from a GRF and parameter estimation are raised.

5.1 MARKOV PROPERTY

The Markov property with respect to a discrete temporal sequence x_0, x_1, $x_2, \ldots$ is defined as [Stark and Woods 2002]

$$\mathrm{f}(x_{i+k}|x_{i-1}, \ldots, x_0) = \mathrm{f}(x_{i+k}|x_{i-1}, \ldots, x_{i-p}) \tag{5.1}$$

for all $k \geq 0$ and for all $i \geq p$. If we consider the case of $k = 0$ and $p = 2$, Equation 5.1 becomes

$$\mathrm{f}(x_i|x_{i-1}, \ldots, x_0) = \mathrm{f}(x_i|x_{i-1}, x_{i-2}). \tag{5.2}$$

This means that the probability of x_i having a certain value given all previous values is the same as the probability of that value given only the previous two values. In other words the value of x_i is only dependent on the two previous temporal values. This property simplifies the evaluation of $\mathrm{f}(x_i|x_{i-1}, \ldots, x_0)$ considerably since only the previous two values have to be investigated.

Translating the Markov property into the n-dimensional spatial domain we get

$$\mathrm{f}(x_i|x_k, \forall k \setminus i) = \mathrm{f}(x_i|x_{\mathcal{N}_i}) \tag{5.3}$$

where x_i is the value of the ith site in an n-dimensional space and $\mathcal{N}_i$ is the set of sites in a neighbourhood of site i. $\forall k \setminus i$ means that k takes all possible values except i. In words Equation 5.3 means that the probability of x_i having a certain value given all other values excluding x_i itself is the same as the probability of x_i having that value given only values in the neighbourhood of x_i.

As an example consider the one-dimensional spatial case with the neighbourhood defined as the two closest neighbours $\mathcal{N}_i = \{i + 1, i - 1\}$. Equation 5.3 becomes

$$\mathrm{f}(x_i|x_k, \forall k \setminus i) = \mathrm{f}(x_i|x_{i+1}, x_{i-1}). \tag{5.4}$$

This means the PDF of x_i is dependent on only its directly neighbouring site values.

5.2 RANDOM FIELD

Let $\mathbf{x}$ be a data array and $\mathbf{i}$ a set of the site locations of that data array so that x_i, $i \in \mathbf{i}$ corresponds to the value of $\mathbf{x}$ at site i of the data array. Let $\mathbf{X}_i$ be a finite space of states of x_i ($x_i \in \mathbf{X}_i$) and $\mathbf{X}$ the finite space of all possible configurations $\mathbf{x} \in \mathbf{X}$ (for example if $\mathbf{X}_{i \in \mathbf{i}} = \{0, 1\}$, corresponding to a binary image, with $\mathbf{i} = \{1, \cdots, n\}$ then $\mathbf{X}$ is of the size 2^n). If $\mathbf{a}$ is a subset of $\mathbf{i}$ ($\mathbf{a} \subset \mathbf{i}$), then $\mathbf{x_a}$ is the data set of subset $\mathbf{a}$ ($\mathbf{x_a} = \{x_i, \forall i \in \mathbf{a}\}$). We now consider the strictly positive PDF $\mathrm{f}(\mathbf{x})_{\mathbf{x} \in \mathbf{X}}$, i.e. $\mathrm{f}(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbf{X}$ and $\sum_{\forall \mathbf{x} \in \mathbf{X}} \mathrm{f}(\mathbf{x}) = 1$. This PDF $\mathrm{f}(\mathbf{x})_{\mathbf{x} \in \mathbf{X}}$ is called a stochastic or random field [Winkler 1995]. Note that $\mathrm{f}(\mathbf{x})$ is the joint density of the data array. In the next section the particular PDF of the Gibbs random field (GRF) is discussed.

5.3 GIBBS RANDOM FIELD

Gibbs random fields (GRF) have been used in statistical mechanics as models for the equilibrium states of large physical systems. Therefore some of the terminology used here reflects this fact.

Consider the PDF of a GRF model for the data $\mathbf{x}$ of the form

$$\mathrm{f}(\mathbf{x}) = \frac{\exp\left[-U(\mathbf{x})\right]}{\sum_{\forall \mathbf{x}' \in \mathbf{X}} \exp\left[-U(\mathbf{x}')\right]} \tag{5.5}$$

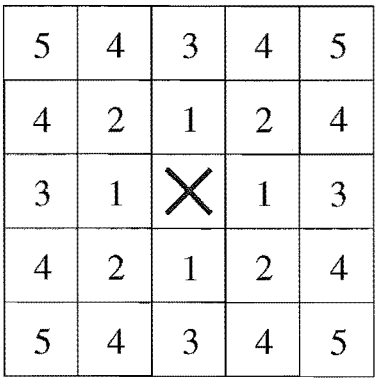


Figure 5.1 The neighbourhood system $\mathcal{N}_i$ in a finite regular lattice system up to the neighbourhood order $D = 5$. The cross indicates the site i and the numbering around it give the order of the neighbourhood. A D order neighbourhood includes all sites from lower order neighbourhoods.

where U is called the energy function. Note that Equation 5.5 is strictly positive and therefore a random field. The denominator of this equation is called the partition function Z_0.

Equation 5.5 returns a probability measure for every configuration $\mathbf{x} \in \mathbf{X}$. It can therefore be used as a stochastic model in a Bayesian framework. In order to ensure that the GRF is a MRF the energy function $U(\mathbf{x})$ has to conform to some rules. For this, the neighbourhood and cliques have to be defined.

5.4 NEIGHBOURHOOD SYSTEM

$\mathcal{N}$ denotes a neighbourhood system for the data. $\mathcal{N}_i \in \mathbf{i}$ is a set of sites which are in the neighbourhood of site i but exclude site i itself ($i \notin \mathcal{N}_i$). The site j is in the neighbourhood of site i ($j \in \mathcal{N}_i$) only if site i is in the neighbourhood of site j ($i \in \mathcal{N}_j$), which enforces a symmetric neighbourhood. Now let $\mathbf{x}_{\mathcal{N}_i}$ denote the set of site values of the neighbourhood of site i. Often neighbourhood systems are in a finite regular lattice system

$$\mathbf{i} = \{(v, w) \in \mathbb{Z} \times \mathbb{Z} : 0 \leq v \leq v_{max}, 0 \leq w \leq w_{max}\} \tag{5.6}$$

as shown in Figure 5.1. The cross in the Figure indicates the site i and the numbering around it gives the order of the neighbourhood D. A D-order neighbourhood includes all sites from lower order neighbourhoods. Therefore the neighbourhood of order $D = 5$ consists of all 24 sites shown in Figure 5.1 excluding the one with the cross. $D = 1$ has the four nearest neighbours as the neighbourhood system (called the four-neighbourhood), while $D = 2$ is also called the eight-neighbourhood. Modifications to the neighbourhood have to be made at the boundary of the finite lattice system. A random field $\mathrm{f}(\mathbf{x})$ is a Markov random field with regard to the neighbourhood system $\mathcal{N}$ if for all $\mathbf{x} \in \mathbf{X}$,

$$\mathrm{f}(x_i | x_j, \forall j \in \{\mathbf{i} \backslash i\}) = \mathrm{f}(x_i | \mathbf{x}_{\mathcal{N}_i}). \tag{5.7}$$

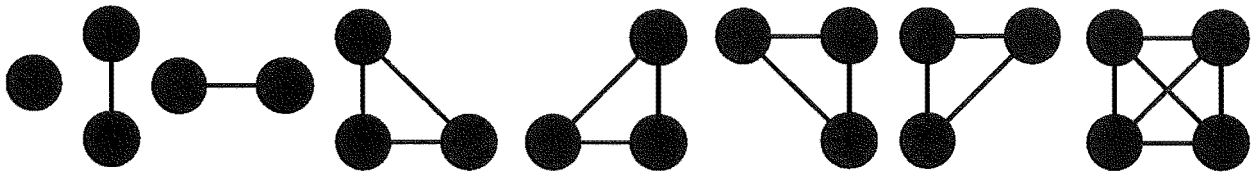


Figure 5.2 All possible clique types of the eight-neighbourhood. The black disks indicate the site locations and sites belonging to one clique are joined by lines. Note that the four-neighbourhood has the first three clique types.

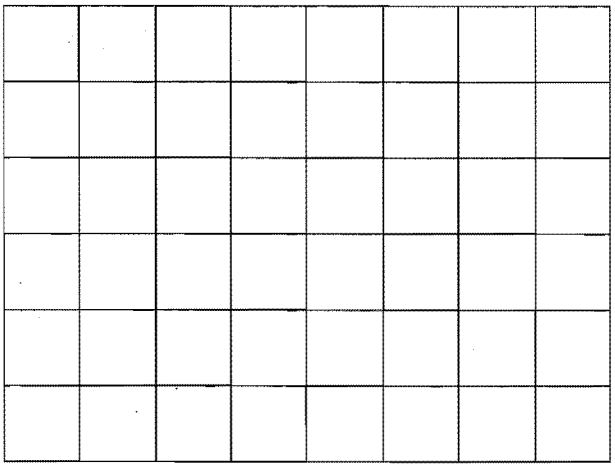


Figure 5.3 A 6×8 site array, each square indicates the location of a site.

5.5 CLIQUES

A clique is a subset of sites in which every site is a neighbour of every other site in the subset. Single sites are also cliques as a special case. The neighbourhood system therefore determines the number of subsets of sites that are cliques. Figure 5.2 shows all possible clique types for the second order neighbourhood system (the eight-neighbourhood) where the black discs indicate the site locations and the sites which belong to one clique are joined by lines. Note that the first three cliques in Figure 5.2 are the clique types for the first order neighbourhood (the four-neighbourhood).

Functions will be defined for the site values of cliques. As each clique type can have its own function, some notation is introduced at this point. Let C be a clique type out of the set $\mathcal{C}$ of all possible clique types defined by the neighbourhood system $\mathcal{N}$. Further, let c denote one of the cliques and $\mathcal{C}_{\mathbf{i}}$ denote the set of all cliques in $\mathbf{i}$. The number of sites within a clique is called the size of the clique (a clique with three sites is therefore called a clique of size three).

In image modelling the selection of clique types used is often limited to clique types of size one and two. This limitation has been found to limit the complexity of the model but still provide a flexible prior model.

As an example, consider a 6×8 site array $\mathbf{i}$ as shown in Figure 5.3 and the four-neighbourhood system with its clique types of size two (see Figure 5.4). Figure 5.5 shows then all possible cliques $c \in \mathcal{C}_{\mathbf{i}}$ of size two for the array.

In the next section the functions on the clique types are introduced. These are called potentials.

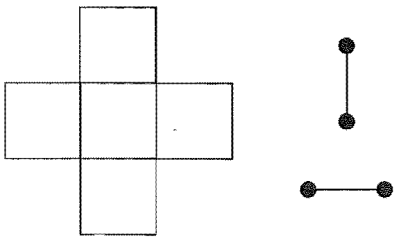


Figure 5.4 The four-neighbourhood system and the corresponding clique types of size two.

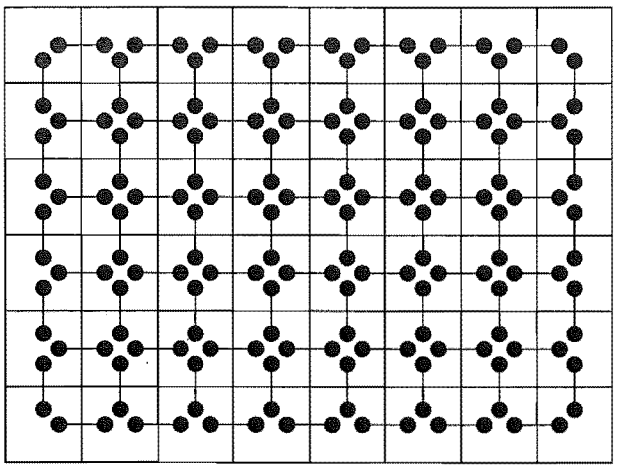


Figure 5.5 All possible cliques of size two for the four-neighbourhood system in the 6×8 site array.

5.6 POTENTIALS

The energy function U in Equation 5.5 can be expressed as a sum of potentials. A potential is a family $\{V_{\mathbf{a}} : \mathbf{a} \subset \mathbf{i}\}$ of functions on $\mathbf{X}$ such that

$$V_{\emptyset}(\mathbf{x}) = 0, \tag{5.8}$$

$$V_{\mathbf{a}}(\mathbf{x}) = V_{\mathbf{a}}(\mathbf{x}') \quad \text{if} \quad \mathbf{x} = \mathbf{x}', \tag{5.9}$$

where $\emptyset$ is the empty set, and

$$U = \sum_{\mathbf{a} \subset \mathbf{i}} V_{\mathbf{a}}. \tag{5.10}$$

Given a neighbourhood system $\mathcal{N}$, a potential V is called a neighbour potential with regard to $\mathcal{N}$ if $V_{\mathbf{a}} = 0$ whenever $\mathbf{a}$ is not a clique. If $V_{\mathbf{a}} = 0$ for all $\mathbf{a}$ of size greater than two ($|\mathbf{a}| > 2$) then V is a pair potential. Note that for every clique type a different potential function can be used. This enables one to enforce higher horizontal site interaction than vertical site interaction for example.

The energy function for the neighbour potential can now be written specifically in terms of cliques

$$U(\mathbf{x}) = \sum_{\forall c \in C_{\mathbf{i}}} V_c(\mathbf{x}), \tag{5.11}$$

and therefore the PDF of the GRF with regard to a neighbourhood system can be written as

$$\Pi(\mathbf{x}) = \frac{1}{Z_0} \exp\left[-\sum_{\forall c \in \mathcal{C}_\mathbf{i}} V_c(\mathbf{x})\right] \tag{5.12}$$

$$Z_0 = \sum_{\forall \mathbf{x}' \in \mathbf{X}} \exp\left[-\sum_{\forall c \in \mathcal{C}_\mathbf{i}} V_c(\mathbf{x}')\right], \tag{5.13}$$

where Z_0 is called the partitioning function. This type of GRF is called a neighbour Gibbs field. Potentials define energy functions and thus random fields.

5.7 GIBBS RANDOM FIELDS AND MARKOV RANDOM FIELDS

Besag [1974] showed that a unique GRF exists for every MRF and vice-versa as long as the GRF is defined in terms of potential functions on cliques as defined in the previous Section. The proof given in Besag [1974] is based on the Hammersley-Clifford theorem which itself was never published by the authors. Another proof based on the Möbius inversion is given in Winkler [1995]. Note however that not every MRF is a GRF and not every GRF is a MRF [Dubes and Jain 1989]. Using cliques that are not based on a neighbourhood system can lead to distributions which do not display the required behaviour. It can therefore be said that defining a GRF model based on cliques derived from a neighbourhood system will yield a MRF.

5.8 EXAMPLE

Consider a 6×8 site array $\mathbf{i}$ of the form shown in Figure 5.3 and a four neighbourhood system $\mathcal{N}$ as shown in Figure 5.4 with its corresponding clique types of size two. Figure 5.5 shows all possible cliques $\mathcal{C}_\mathbf{i}$ in the site array $\mathbf{i}$. In this example only the potential functions for cliques with size two are not equal to zero and defined by

$$V_c(x_i, x_j) = \begin{cases} -\beta & \text{if } x_i = x_j, \\ +\beta & \text{if } x_i \neq x_j. \end{cases} \qquad c \in \mathcal{C}_\mathbf{i} \tag{5.14}$$

This means that if neighbouring data values are equal the potential function returns a negative value $-\beta$ and if they are unequal a positive value β is returned. Note that $V_c(x_i, x_j) = V_c(x_j, x_i)$. The Gibbs density for the MRF can therefore be written as

$$\Pi(\mathbf{x}) = \frac{1}{Z_0} \exp\left[-\sum_{c \in \mathcal{C}_\mathbf{i}} V_c(x_i, x_j)\right] \tag{5.15}$$

where

$$Z_0 = \sum_{\forall \mathbf{x}' \in \mathbf{X}} \exp\left[-\sum_{c \in \mathcal{C}_\mathbf{i}} V_c(\mathbf{x}')\right]. \tag{5.16}$$

In the special case of $\mathbf{X}_{i\in\mathbf{i}} = \{0,1\}$ this MRF is called an Ising model [Winkler 1995]. It is clear that for $\beta > 0$ the maximum of $\Pi(\mathbf{x})$ lies at the x where every x_i has the same value (like a single colour image). If $\mathbf{X}_{i\in\mathbf{i}} = \{0,\ldots,9\}$ (ten different data values) and if the size of $\mathbf{i}$ equals $6 \times 8 = 48$, then $\mathbf{X}$ is of the size 10^{48}. This means that 10^{48} different x configurations are possible. It thus becomes impractical to directly compute Z_0 for any realistic sized image. In Section 5.10 methods for estimating the Gibbs densities are discussed.

In order to show that the spatial Markov property of Equation 5.7 holds for the above example, the Bayesian theorem in Equation 3.8 can be invoked to write the probability of a data site having a value x_k given all other data site values, thus

$$\mathrm{f}(x_k|\{\mathbf{x} \setminus x_k\}) = \frac{\mathrm{f}(\mathbf{x})}{\mathrm{f}(\mathbf{x} \setminus x_k)}. \tag{5.17}$$

The set of cliques $C_\mathbf{i}$ in the data array can be split into two sets, one that is the set of all cliques which do not contain the site k denoted $C_{\mathbf{i}\setminus k}$ and the other is the set of the remaining cliques which contain the site k denoted C_k. Note that the set of sites covered by C_k is $\{k, \mathcal{N}_k\}$. With this, Equation 5.15 is written as

$$\Pi(\mathbf{x}) = \frac{\exp\left[-\sum_{\forall c\in C_\mathbf{i}} V_c(x_i,x_j)\right]}{Z_0} = \frac{\exp\left[-\sum_{\forall c\in C_{\mathbf{i}\setminus k}} V_c(x_i,x_j)\right]\exp\left[-\sum_{\forall c\in C_k} V_c(x_k,x_j)\right]}{Z_0}. \tag{5.18}$$

Using the discrete version of the total probability definition in Equation 3.12,

$$\mathrm{f}(\mathbf{x} \setminus x_k) = \sum_{\forall x'_k\in\mathbf{X}_k} \mathrm{f}(\mathbf{x}) = \sum_{\forall x'_k\in\mathbf{X}_k} \frac{\exp\left[-\sum_{\forall c\in C_{\mathbf{i}\setminus k}} V_c(x_i,x_j)\right]\exp\left[-\sum_{\forall c\in C_k} V_c(x'_k,x_j)\right]}{Z_0} \tag{5.19}$$

$$= \frac{\exp\left[-\sum_{\forall c\in C_{\mathbf{i}\setminus k}} V_c(x_i,x_j)\right]\sum_{\forall x'_k\in\mathbf{X}_k}\exp\left[-\sum_{\forall c\in C_k} V_c(x'_k,x_j)\right]}{Z_0}. \tag{5.20}$$

Note that all cliques affected by x'_k are in the set of cliques C_k. Equation 5.17 becomes

$$\mathrm{f}(x_k|\{\mathbf{x} \setminus x_k\}) = \frac{\exp\left[-\sum_{\forall c\in C_{\mathbf{i}\setminus k}} V_c(x_i,x_j)\right]\exp\left[-\sum_{\forall c\in C_k} V_c(x_k,x_j)\right]}{\exp\left[-\sum_{\forall c\in C_{\mathbf{i}\setminus k}} V_c(x_i,x_j)\right]\sum_{\forall x'_k\in\mathbf{X}_i}\exp\left[-\sum_{\forall c\in C_k} V_c(x'_k,x_j)\right]} \tag{5.21}$$

$$= \frac{\exp\left[-\sum_{\forall c\in C_k} V_c(x_k,x_j)\right]}{\sum_{\forall x'_k\in\mathbf{X}_i}\exp\left[-\sum_{\forall c\in C_k} V_c(x'_k,x_j)\right]} = \mathrm{f}(x_k|\mathbf{x}_{\mathcal{N}_k}). \tag{5.22}$$

This proves that the spatial Markov property as given in Equation 5.7 applies to this example.

5.9 MORE ON THE GRF MODEL AND POTENTIAL FUNCTIONS

The main advantage of the GRF model is the relationship between the joint density as given in Equation 5.12 and the local conditional density $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$. While the joint density is impractical

to evaluate for any reasonably sized image the local conditional density is fast to compute and on a local basis holds all the characteristics of the joint density. The general form of the local conditional PDF $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$ of the GRF $\Pi(\mathbf{x})$ can easily be derived following the example given in Section 5.8 to

$$\Pi(x_i|\mathbf{x}_{\mathcal{N}_i}) = \frac{1}{Z}\exp\left[-U(x_i, \mathbf{x}_{\mathcal{N}_i})\right] = \frac{1}{Z}\exp\left[-\sum_{\forall c \in \mathcal{C}_i} V_c(\mathbf{x})\right], \tag{5.23}$$

where

$$Z = \sum_{\forall x_i' \in \mathbf{X}_i} \exp\left[-\sum_{\forall c \in \mathcal{C}_i} V_c(\mathbf{x}')\right], \tag{5.24}$$

and $V_c(\mathbf{x}')$ denotes the potential for clique c where the site value x_i is replaced with x_i'. Instead of the probability of a single site given its neighbourhood, the probability of a selection of sites given its neighbourhood can be computed in the same fashion.

The potential functions on cliques used in the remainder of this thesis are defined as follows: The single site clique potential is assigned a value α depending on the site value.

$$V_C(\mathbf{x}) = \alpha(\mathbf{x}) \quad \text{for} \quad |c| = 1. \tag{5.25}$$

Multiple site cliques are assigned a β value with a negative sign if all the sites have equal values and a positive sign otherwise:

$$V_c(\mathbf{x}) = \begin{cases} -\beta_C & \text{if all } x \in \mathbf{x}_c \text{ are equal} \\ +\beta_C & \text{otherwise} \end{cases} \qquad c \in C \in \mathcal{C}, |c| \geq 2. \tag{5.26}$$

Note that each individual clique type may be assigned a different β value according to Equation 5.26. This model is the same as the one used by Geman and Geman [1984] and Derin and Elliott [1987] and most other discrete GRF applications. In most cases reported the models used have been restricted to cliques of size one and two ($|c| \leq 2$) and a small neighbourhood (mainly the four- and eight-neighbourhood). Larger neighbourhoods, the use of which would result in more clique types, are mostly employed in GRF models for textures. Dependencies between label values in the lattice system are introduced by a larger neighbourhood over a wider area making the model more restrictive. Interestingly, in Chapter 10 it is shown that in some cases the use of a four-neighbourhood as a prior model can produce better results than can an eight-neighbourhood.

The two main problems with GRF models are the difficulty of sampling from their distribution and the difficulty of estimating model parameters. Both difficulties arise because the partitioning function is impractical to compute for any practical size of image. Both problems are often solved by using the local properties (local conditional PDF). Because of the spatial Markov property, the local conditional PDF is dependent only on the site under investigation and its neighbourhood. The next two sections describe techniques for sampling from a GRF and for estimating model parameters for a GRF.

5.10 SAMPLING FROM A GRF

Directly sampling (creating a realisation) from a GRF with given model parameters is impractical due to the intractability of the partition function Z_0. A practical alternative is to perform local sampling with a rejection or acceptance rule in an iterative sweeping fashion. Two commonly used methods of sampling from a GRF are given here.

5.10.1 Metropolis sampler

The first one is called the Metropolis sampler, described by Chen [1988]; Dubes and Jain [1989] as:

1. Start with a random initial configuration $\mathbf{x} \in \mathbf{X}$.
2. Sweep through $\mathbf{i}$, covering all locations. At every location i:
   (a) Choose a random data value for location i ($x_i' \in \mathbf{X}_i$) and call the new configuration $\mathbf{x}'$ ($\mathbf{x}'$ only differs at one point from $\mathbf{x}$).
   (b) Let $p = \min[1, \Pi(\mathbf{x})/\Pi(\mathbf{x}')]$ (see Equation 5.12 for $\Pi(\cdot)$).
   (c) Replace $\mathbf{x}$ with $\mathbf{x}'$ with probability p.
3. Repeat Step 2 for a set number of iterations or until a stopping criterion is met.

Note that if $\mathbf{x}'$ only differs from $\mathbf{x}$ at site i then

$$\frac{\Pi(\mathbf{x})}{\Pi(\mathbf{x}')} = \frac{\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})}{\Pi(x_i'|\mathbf{x}_{\mathcal{N}_i})} = \frac{\exp\left[-U(x_i, \mathbf{x}_{\mathcal{N}_i})\right]}{\exp\left[-U(x_i', \mathbf{x}_{\mathcal{N}_i})\right]} = \exp\Big[-\left(U(x_i, \mathbf{x}_{\mathcal{N}_i}) - U(x_i', \mathbf{x}_{\mathcal{N}_i})\right)\Big] \tag{5.27}$$

which is fast to compute for small neighbourhoods.

5.10.2 Gibbs sampler

Geman and Geman [1984] introduced the Gibbs sampler, which is sweeping through $\mathbf{i}$ and replacing every x_i with a sample from the local conditional PDF in an iterative fashion. The Gibbs sampler method is:

1. Start with a random initial configuration $\mathbf{x} \in \mathbf{X}$.
2. Sweep through $\mathbf{i}$, covering all locations. At every location i:
   (a) Choose x_i randomly but with the probability $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$ (for this, $\Pi(x_i'|\mathbf{x}_{\mathcal{N}_i})$ has to be evaluated for all $x_i' \in \mathbf{X}_i$).
3. Repeat Step 2 for a set number of iterations or until a stopping criteria is met.

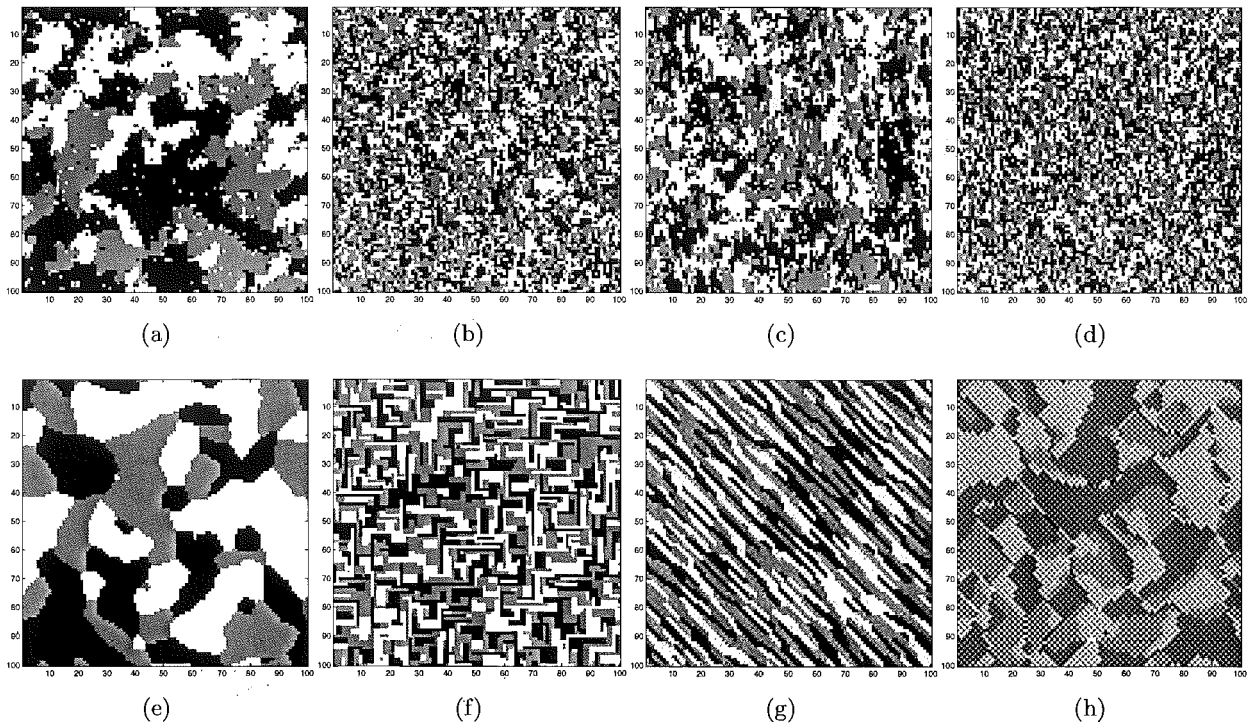


Figure 5.6 Realisations of a GRF with the potential function given in Equation 5.26. Images are of the size 100 × 100 pixels and contain 3 classes. Fifty iterations (complete sweeps) where used to generate these with the Metropolis sampler. (a) – (d) are based on the four-neighbourhood and (e) – (h) on the eight-neighbourhood. The parameter vectors for the cliques are as follows: (a) $\Theta = [0.6\ \ 0.6]$, (b) $\Theta = [0.3\ \ 0.3]$, (c) $\Theta = [0.6\ \ 0.3]$, (d) $\Theta = [0.3\ \ 0.15]$, (e) $\Theta = [2\ \ 2\ \ 1\ \ 1]$, (f) $\Theta = [2\ \ 2\ \ -1\ \ -1]$, (g) $\Theta = [2\ \ 2\ \ 2\ \ -2]$ and (h) $\Theta = [-2\ \ -2\ \ 4\ \ 4]$.

5.10.3 Comparison and examples

Both methods have been shown to converge to a sample from $\Pi(\mathbf{x})$ given in Equation 5.12 starting at any initial configuration. It was found that a random initial configuration (uniformly sampled local site values from $x_i \in \mathbf{X}_i$) converges in a reasonable amount of time [Dubes and Jain 1989]. The second method, the Gibbs sampler, is computationally more expensive than the first method, since k (k being the number of elements in $\mathbf{X}_i$) exponential functions have to be evaluated in order to sample from the local conditional PDF $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$ rather than only one as part of $\min[1, \Pi(\mathbf{x})/\Pi(\mathbf{x}')]$.

Figure 5.6 shows samples from a GRF generated by the Metropolis sampler method presented in Chen [1988] and Dubes and Jain [1989]. The sample images are 100 ×100 pixels with three classes. Fifty iterations (complete sweeps through the whole image) were used. Chen [1988] described the convergence of the sampling method used and found that 50 iterations were sufficient to produce a good sample for the given parameters. The Figures (a) – (d) are based on the four-neighbourhood and the Figures (e) – (h) on the eight-neighbourhood. Only cliques of size two were used and the potential model is given in Equation 5.26. The parameter vector for the four-neighbourhood was given by $\Theta = [\ |\ \ -]$ and for the eight-neighbourhood by $\Theta = [\ |\ \ -\ \ \backslash\ \ /\ ]$ where the lines indicate the clique type (for example | is the vertical clique of size two shown in Figure 5.2 second

from left). For example $\Theta = [0.6 \;\; 0.3]$ means that the β value for the vertical clique is $\beta_| = 0.6$ and for the horizontal clique $\beta_- = 0.3$.

The figures show that higher β values produce higher bonds between the pixels of the corresponding cliques. Figure 5.6(a) consists of larger areas than Figure 5.6(b) where the β values are halved. In Figures 5.6(c) and (d) more vertical than horizontal features are present since $\beta_| > \beta_-$. For the examples with the eight-neighbourhood (Figures 5.6(e) to (h)) the effects are more complicated due to the higher complexity of the model. Figure 5.6(g) is a good example in which only the $/$ cliques are penalised and therefore the features in the $\backslash$ direction dominate.

5.11 PARAMETER ESTIMATION

The GRF, if used as a prior model in a classification, describes the resulting label image. The GRF prior model will make some label images more likely than others and therefore guide the classification. For example if the prior knowledge exists that the scene under investigation is made up of large horizontal features the GRF model parameters have to be chosen accordingly ($\beta_| < \beta_-$). Therefore having a good method of estimating the GRF model parameters from a sample label image will yield a better GRF prior model and hence better classification results.

The parameter estimation problem can be defined as follows. For a given sample image, estimate the GRF model parameters. Let Θ be the set of parameters of the GRF (in the example (Section 5.8) $\Theta = \beta$). The task now is to obtain an estimate $\hat{\Theta}$ of Θ for a given sample image $\mathbf{x}$ and a specific GRF model $\Pi(\mathbf{x})$. A traditional way of obtaining $\hat{\Theta}$ would be the maximum likelihood estimate

$$\hat{\Theta} = \arg\max_{\Theta} \Pi(\mathbf{x}, \Theta). \tag{5.28}$$

However, this method cannot be used due to the difficulties in evaluating the partition function Z_0 of a specific $\Pi(\mathbf{x})$ (Equation 5.12). Chen [1988] and Dubes and Jain [1989] discuss several different methods of parameter estimations. The most commonly used is the coding method developed by Besag described below. In Chapter 8 an alternate method of parameter estimation is presented.

5.11.1 Coding method

Besag [1974] developed the coding method in which the site array $\mathbf{i}$ is partitioned into conditionally independent sets of samples each called a "code". This means that a specific site i in a code has no other site in that same code as a neighbour. For each neighbourhood system there exists an optimal way of partitioning the site array in order to maximise the number of samples sites in each code.

Let $\mathbf{i}_1$ denote the set of sites of the first code and so on. To maximise the number of samples in one code the sites are selected as closely to each other as possible without being in each others neighbourhood. Figures 5.7 and 5.8 show the site array $\mathbf{i}$ with numbers indicating which code each site belongs to for the four-neighbourhood and the eight-neighbourhood respectively.

$$\begin{array}{cccccccccc} 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & \cdots \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & \cdots \\ 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & \cdots \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & \cdots \\ 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & \cdots \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Figure 5.7 Site array $\mathbf{i}$ with numbers at each site indicating which code the site belongs to for the four-neighbourhood.

$$\begin{array}{cccccccccc} 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & \cdots \\ 3 & 4 & 3 & 4 & 3 & 4 & 3 & 4 & 3 & 4 & \cdots \\ 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & \cdots \\ 3 & 4 & 3 & 4 & 3 & 4 & 3 & 4 & 3 & 4 & \cdots \\ 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 2 & \cdots \\ 3 & 4 & 3 & 4 & 3 & 4 & 3 & 4 & 3 & 4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Figure 5.8 Site array $\mathbf{i}$ with numbers at each site indicating which code the site belongs to for the eight-neighbourhood.

It is easy to show that these sites are conditionally independent by using the Bayes formula. If site i and site j belong to the same code k ($\{i,j\} \in \mathbf{i}_k$) then it is trivial to show using Equation 5.23 that the following is valid

$$\Pi(x_i, x_j|\mathbf{x}_{\mathcal{N}_i}, \mathbf{x}_{\mathcal{N}_j}) = \Pi(x_i|\mathbf{x}_{\mathcal{N}_i})\Pi(x_j|\mathbf{x}_{\mathcal{N}_j}). \tag{5.29}$$

The parameter estimate $\hat{\Theta}_k$ is then obtained by maximising the likelihood (or log likelihood) of the code k

$$\hat{\Theta}_k = \arg\max_{\Theta} \prod_{\forall i \in \mathbf{i}_k} \Pi(x_i|\mathbf{x}_{\mathcal{N}_i}, \Theta). \tag{5.30}$$

Note that for the four- and eight-neighbourhood different numbers of estimates $\hat{\Theta}_k$, two and four respectively, can be obtained. This depends on the number of codes available. These can be averaged in order to obtain an improved parameter estimate

$$\hat{\Theta} = \frac{1}{K}\sum_{k=1}^{K} \hat{\Theta}_k \tag{5.31}$$

where K is the total number of codes available. The optimisation in Equation 5.30 can be performed by a straight forward gradient based method.

In the next chapter the MAP approach to classification of SAR images will be discussed, implementing the speckle models and GRF model introduced in this and the previous chapter.

Chapter 6

MAP APPROACH TO CLASSIFICATION OF SAR DATA

In this chapter the results from the two previous chapters are combined to give a MAP approach of SAR data analysis. Forming a MAP estimate (see Section 3.2) requires the likelihood function and the prior model. In the case discussed here the likelihood function describes the connection between the true ground cover and the measured data in the form of a PDF. Such functions for SAR have been described in Chapter 4. The prior model describes what is known *a priori* about the result, being the classified SAR data. A commonly used prior model is the GRF discussed in Chapter 5. This prior model introduces the dependency of a pixel on its neighbouring pixels and provides a tool of adjusting this dependency by means of the model parameters.

The MAP estimate of the classified label image $\hat{\mathbf{r}}$ is defined as

$$\hat{\mathbf{r}} = \arg\max_{\mathbf{r}} \left[f(\mathbf{s}|\mathbf{r}) \Pi(\mathbf{r}) \right], \tag{6.1}$$

where $\mathbf{s}$ is the measured SAR data (known) and $\mathbf{r}$ is the true (unknown) label image (each pixel has a label value corresponding to a ground cover class such as "bare ground" or "water" for example). The likelihood function $f(\mathbf{s}|\mathbf{r})$ in Equation 6.1 is the appropriate speckle model for the type of SAR data that has been measured (see Chapter 4). In addition to the speckle model, texture models can be introduced at this point in the form of an additional PDF. Each ground cover class has a set of parameters (for example the mean and variance of the intensity response from such a ground cover) for the likelihood function. Typically these parameters are obtained from a training SAR data set where the ground cover is known beforehand. Obtaining parameters from a data set in this way is called supervised training. Unsupervised training in this context would be estimating the parameters from a data set where the ground cover is not known. In most SAR classifications, however, unsupervised training of the parameters on a pixel-by-pixel basis is impractical due to the severity of the speckle noise present.

The prior model, $\Pi(\mathbf{r})$, is a PDF only for the resulting label image. This model determines which realisations of $\mathbf{r}$ are more likely than others. Here the GRF is used as a prior model (see Chapter 5) and the four- and eight-neighbourhood are the most commonly used neighbourhood size in SAR data classification. The parameters of the prior model can be estimated before hand from a training set (see Section 5.11) or on the fly [Lakshmanan and Derin 1989].

Equation 6.1 presents a difficult optimisation problem. The function is typically non-convex and the search space for $\mathbf{r}$ can be very large. As an example take an 1000×1000 pixel image with 10 classes: for this there are 1000000 parameters and $10^{1000000}$ different configurations of $\mathbf{r}$ to be considered. The two most commonly used methods for this optimisation are presented here. The first is based on a simulated annealing technique [Geman and Geman 1984] while the second is called iterated conditional modes (ICM) [Besag 1986]. The work done in this thesis is based on the simulated annealing method and the ICM method is only presented for completeness. For a short overview of these methods the reader is referred to Dubes and Jain [1989] while Winkler [1995] gives a thorough insight into the algorithms including proofs.

6.1 SIMULATED ANNEALING

Simulated annealing is derived from the physical annealing of, for example, metals or glass. A metal is heated to a very high temperature where the molecules are completely randomly distributed and moving around at random (a high energy state). The temperature is then slowly lowered following a predetermined temperature profile. The random movement of the molecules decreases as the temperature decreases. If the temperature is decreased slowly enough the molecules settle in their lowest energy state which in many cases is a mono-crystal type configuration (a low energy state).

Simulated annealing is a stochastic optimisation method. Given an energy function $E(\mathbf{x})$, an $\mathbf{x}$ is to be derived that minimises the energy function or at least gets close to the global minimum. For this purpose the energy function is written as part of the Boltzmann distribution:

$$\mathrm{f}(\mathbf{x}) = \frac{1}{Z_\mathrm{B}} \exp\left[-\frac{1}{k_\mathrm{B} T_\mathrm{A}} E(\mathbf{x})\right], \tag{6.2}$$

where T_A is called the annealing temperature, k_B the Boltzmann's constant and Z_B is a normalising constant (note that Z_B is a function of T_A). Note that the Boltzmann distribution is flat (close to uniform) for high values of T_A and steep for low values of T_A.

The simulated annealing algorithm is started with an initial estimate of $\mathbf{x}$, which can be completely random. This configuration is then randomly changed over many iterations. If the changed configuration of $\mathbf{x}$ is called $\mathbf{x}'$, the change is accepted or rejected depending on $\mathrm{f}(\mathbf{x})$ and $\mathrm{f}(\mathbf{x}')$. With each iteration the temperature T_A is lowered following a predetermined temperature profile. It is important that a change can be accepted even if the energy of the changed configuration is higher than the previous configuration. This feature enables simulated annealing to escape from being trapped in a local minimum.

Winkler [1995] gives the proof that a cooling schedule exists such that the above algorithm reaches the global minimum. This however requires that the temperature T_A is decreased very slowly over many iterations which makes this algorithm computationally expensive and a slow optimisation method. In practice a faster than ideal cooling schedule is used; the global minimum might not be reached but generally a good estimate of the minimum can be obtained.

In the case of image classification as discussed here the energy function is derived from the *a posteriori* probability

$$\mathrm{f}(\mathbf{r}|\mathbf{s}) \propto \mathrm{f}(\mathbf{s}|\mathbf{r})\Pi(\mathbf{r}) \tag{6.3}$$

where $\mathbf{r}$ is the resulting label image, $\mathbf{s}$ the supplied SAR data and $A \propto B$ means A is proportional to B. The energy function is then given as

$$E(\mathbf{r},\mathbf{s}) = \ln\left[\mathrm{f}(\mathbf{s}|\mathbf{r})\Pi(\mathbf{r})\right]. \tag{6.4}$$

The Boltzmann distribution then becomes

$$\mathrm{f}(\mathbf{r},\mathbf{s}) = \frac{1}{Z_\mathrm{B}} \exp\left[-\frac{1}{T_\mathrm{A}} E(\mathbf{r},\mathbf{s})\right], \tag{6.5}$$

where the Boltzmann constant has been omitted for convenience.

The Metropolis sampler is used to change $\mathbf{r}$ and determine whether the change should be accepted or not (see Section 5.10.1). The simulated annealing algorithm as implemented here is:

1. Start with a random initial configuration $\mathbf{r}$ and T_A with a sufficiently high value.
2. Sweep through $\mathbf{i}$, covering all locations. At every location i:
   (a) Choose a random label value r_i' for location i and call the new configuration $\mathbf{r}'$ ($\mathbf{r}'$ only differs at one point from $\mathbf{r}$).
   (b) Let $p = \min[1, \mathrm{f}(\mathbf{r}',\mathbf{s})/\mathrm{f}(\mathbf{r},\mathbf{s})]$ where $\mathrm{f}(\mathbf{r},\mathbf{s})$ is the Boltzmann density in Equation 6.5.
   (c) Replace $\mathbf{r}$ with $\mathbf{r}'$ with the probability p.
3. Lower T_A after each sweep.
4. Repeat Step 2 for a set number of iterations or until a stopping criterion is met.

Note that, as in the Metropolis sampler case, the term $\mathrm{f}(\mathbf{r},\mathbf{s})/\mathrm{f}(\mathbf{r}',\mathbf{s})$ can be reduced to the local term properties

$$\frac{\mathrm{f}(\mathbf{r},\mathbf{s})}{\mathrm{f}(\mathbf{r}',\mathbf{s})} = \frac{\exp\left[-\frac{1}{T_\mathrm{A}} E(\mathbf{r},\mathbf{s})\right]}{\exp\left[-\frac{1}{T_\mathrm{A}} E(\mathbf{r}',\mathbf{s})\right]} = \left[\frac{\mathrm{f}(\mathbf{s}|\mathbf{r})}{\mathrm{f}(\mathbf{s}|\mathbf{r}')}\frac{\Pi(\mathbf{r})}{\Pi(\mathbf{r}')}\right]^{-\frac{1}{T_\mathrm{A}}} = \left[\frac{\mathrm{f}(s_i|r_i)}{\mathrm{f}(s_i|r_i')}\frac{\Pi(r_i|\mathbf{r}_{\mathcal{N}_i})}{\Pi(r_i'|\mathbf{r}_{\mathcal{N}_i})}\right]^{-\frac{1}{T_\mathrm{A}}} \tag{6.6}$$

which is easy and fast to compute.

Let ι be the iteration counter (after each complete sweep through $\mathbf{i}$, ι is increased by one). A commonly used temperature schedule for this type of simulated annealing is

$$T_\mathrm{A} = \frac{C_\mathrm{A}}{\log(\iota)} \tag{6.7}$$

where C_A is called the cooling constant. For a high value of T_A nearly every random change of the configuration will be accepted and the above algorithm will produce a completely random $\mathbf{r}$ which

is independent from $\mathbf{r}$ of the previous iteration. A very small value for T_A will force $\mathbf{r}$ into a local minimum after a few iterations close to the initial $\mathbf{r}$ configuration.

The starting temperature has to be chosen high enough so that nearly all new configurations are accepted. This means that the result is independent of the initial starting configuration of $\mathbf{r}$. The stopping criterion can have many forms. One way to determine when the iterations should be stopped is to check if there are no (or only a small number of) changes accepted for a certain number of iterations. This means that the algorithm has converged to a configuration close to a minimum. Another stopping criterion is to stop when a certain temperature is reached (which is the same as limiting the number of iterations, if the cooling schedule has been defined). In order to speed up the simulated annealing process, different visiting schemes have been investigated. It has been shown that a random visiting scheme of pixel sites performs better than a raster sweeping scheme. Note that the algorithm can be implemented in a parallel fashion which would decrease the computing time. For a detailed discussion of convergence, speeding up the simulated annealing, start temperature, stopping criteria, visiting scheme and parallel computing the reader is referred to Winkler [1995].

6.2 ITERATED CONDITIONAL MODES

The iterated conditional modes (ICM) method was developed by Besag [1986] and maximises in an iterative fashion the local conditional probability of all sites, i.e. in a "conditional mode". It can be interpreted as a special case of simulated annealing with $T_A \rightarrow 0$. The metropolis sampler is in effect replaced by a greedy optimisation algorithm and will approach a local maximum in a few iterations. To ensure that the local maximum is close to the global maximum, the maximum likelihood estimate is used as the initial starting configuration. The maximum likelihood can easily be obtained for most SAR models.

The temperature is omitted and the Boltzmann distribution becomes a Gibbs distribution with regard to the energy function (defined in Equation 6.4)

$$\mathrm{f}(\mathbf{r},\mathbf{s}) = \frac{1}{Z_I} \exp\big[-E(\mathbf{r},\mathbf{s})\big], \tag{6.8}$$

where Z_I is a normalising constant. The structure of the ICM algorithm is as follows:

1. Start with the maximum likelihood configuration $\mathbf{r}_{\text{initial}} = \max_{\mathbf{r}} \mathrm{f}(\mathbf{s}|\mathbf{r})$ as the initial configuration.

2. Sweep through $\mathbf{i}$, covering all locations. At every location i:

   (a) Choose a random data value r_i' for location i and call the new configuration $\mathbf{r}'$ ($\mathbf{r}'$ only differs at one point from $\mathbf{r}$).

   (b) Replace $\mathbf{r}$ with $\mathbf{r}'$ if $\mathrm{f}(\mathbf{r}',\mathbf{s})/\mathrm{f}(\mathbf{r},\mathbf{s}) > 1$ where $\mathrm{f}(\mathbf{r},\mathbf{s})$ is defined in Equation 6.8.

3. Repeat Step 2 for set number of iterations or until a stopping criterion is met.

The advantage of the ICM method is the fast convergence to a maximum (in most cases only 5 or 6 iterations are needed) and therefore a result is obtained in a very short time. The drawbacks are the fact that the result is very dependent on the Gibbs field in question, the initial configuration and the visiting scheme. For a detailed analysis of ICM the reader is referred to Besag [1986] and Winkler [1995].

Simulated annealing offers a versatile optimisation tool with the option of obtaining a better result by increasing computing time. For this reason simulated annealing is used throughout the thesis as the optimisation tool.

Part II

Contribution: Theory

Chapter 7

NEW LIKELIHOOD FUNCTION FOR POLARIMETRIC SAR DATA

In this chapter the joint density for the three intensities measured by an n-look monostatic polarimetric SAR system is derived. An expression for the intensity statistics of a polarimetric SAR system is required for various kinds of SAR data processing such as maximum likelihood or MAP image estimation [Kay 1993] (see Section 3.2 on page 18 and Chapter 6 for information on the MAP estimate in image processing and Chapter 4 for PDFs of other SAR imaging modes). Such processing is essential for image analysis of remotely sensed data such as speckle noise filtering, image segmentation and classification [Lee et al. 1991, 2001; Lopès and Séry 1997; Rignot and Chellapa 1992] (see Chapter 6 for details). A SAR imaging satellite planned for launch in early 2004 (RADARSAT-2) will, in one imaging mode, provide the averaged intensity values of the three polarisations rather than the full complex covariance matrix, thereby saving bandwidth in the data transfer [Staples and Hornsby 2002].

In Section 7.1 the PDF for the three intensities measured by an n-look monostatic polarimetric SAR system is derived. It has the form of a correlated three-dimensional chi-square distribution. The joint density coincides with that for the average of squares of n samples from a correlated, three-dimensional, complex, circular Gaussian distribution. The expression derived here for the density is in the form of an infinite series that converges rapidly and is simple and fast to compute. This expression will be referred to as the "new likelihood function" throughout this thesis.

In Section 7.2 the new likelihood function is shown to reduce to known forms for three special cases. In the first case the three-polarisation SAR signals are mutually uncorrelated. In the second case one of the three-polarisation SAR signals is uncorrelated to the remaining two. In the third case the correlations between the real and imaginary parts of the complex polarisation SAR signals are removed. For these three cases results have been published and are in agreement with the new likelihood function.

In Section 7.3 a different approach to calculating the new likelihood function is presented. This is done via the characteristic function (or the moment generating function). This approach is pursued here and an alternative form of the joint density is obtained.

In the next Section 7.4 expressions are derived that allow parameters of the new likelihood function to be calculated from samples of intensity data using results from the previous section on

the characteristic function.

In Section 7.5 the conclusions of this Chapter are given. [Let $f(I_X, I_Y, I_Z)$ be the PDF of the three intensity values of n-look monostatic polarimetric SAR data (the new likelihood function), where I_X, I_Y and I_Z correspond to the intensity values from the three polarisations (HH, VV and HV) and are defined as in Equation 4.6 on page 25. In the next section this PDF will be derived.]

7.1 DERIVATION OF THE JOINT DENSITY

For monostatic polarimetric SAR data three complex amplitudes are measured, one for each polarisation response (HH, VV and HV). Let $\tilde{\mathbf{W}} = [\tilde{X}, \tilde{Y}, \tilde{Z}]^T$ where the elements are the three complex amplitudes respectively. The complex multilook covariance matrix $\tilde{\mathbf{R}}$ is then defined as given in Equation 4.18 on page 28. Note that the diagonal elements of the 3×3 matrix $\tilde{\mathbf{R}}$ correspond to the intensity values, I_X, I_Y and I_Z of the three complex amplitudes. The PDF for $\tilde{\mathbf{R}}$, the complex covariance matrix of polarimetric n-look SAR data, is known to be the complex Whishart distribution as given in Equation 4.19 (see Section 4.1.4.1 on page 28).

One way to obtain the PDF of the three intensities, $f(I_X, I_Y, I_Z)$, would be to integrate the complex Wishart PDF, $f(\tilde{\mathbf{R}})$, over the complex off diagonal elements in $\tilde{\mathbf{R}}$. Note that there would be three complex or six real integrals to solve. This approach was considered too difficult. For two polarisations, however, this method has been successfully applied by Lee et al. [1994b] (see Section 4.1.4.2 on page 28).

The approach taken in this thesis to calculate $f(I_X, I_Y, I_Z)$ is based upon the fact that the speckle noise statistics can be described as sampling from a circular Gaussian density. The intensities I_X, I_Y and I_Z of n-look monostatic polarimetric SAR data can be seen as the intensities from n averaged samples from a three-dimensional zero-mean complex circular Gaussian PDF. The three-dimensional zero-mean complex circular Gaussian PDF is rewritten as a six-dimensional zero-mean real Gaussian PDF (splitting the complex terms into real and imaginary parts) with the covariance matrix reflecting the complex circular Gaussian properties. Integrating this PDF over three n-dimensional spheres of constant radii, the joint PDF for the amplitudes is obtained which is straightforwardly transformed into the desired PDF for intensities. To evaluate the integrals, results from Miller et al. [1958] on the distribution of the sum of independent samples from a real correlated Gaussian distribution are used. The complex and real zero-mean circular Gaussian densities used here were introduced in Section 4.1.4 on page 26, but some of the notation will be repeated here for convenience.

Let $[\tilde{X}_i, \tilde{Y}_i, \tilde{Z}_i]^T$ be the ith sample from a three-dimensional complex zero-mean circular Gaussian random process, where

$$\begin{aligned} \tilde{X}_i &= X_{Ri} + jX_{Ii} \\ \tilde{Y}_i &= Y_{Ri} + jY_{Ii} \\ \tilde{Z}_i &= Z_{Ri} + jZ_{Ii}. \end{aligned} \tag{7.1}$$

The corresponding vector of sample scalars is denoted by

$$\mathbf{W}_i = [X_{\mathrm{R}i}, X_{\mathrm{I}i}, Y_{\mathrm{R}i}, Y_{\mathrm{I}i}, Z_{\mathrm{R}i}, Z_{\mathrm{I}i}]^T \tag{7.2}$$

where $\cdot^T$ denotes transposition. Note that the different samples are independent from each other ($\mathbf{W}_i$ is independent of $\mathbf{W}_k$ for $i \neq k$). The covariance matrix $\mathbf{C} = \mathrm{E}[\mathbf{W}_i \mathbf{W}_i^T]$ takes the form [Goodman 1963]

$$\mathbf{C} = \left[\begin{array}{cc|cc|cc} \sigma_X^2 & 0 & c_{XYA} & -c_{XYC} & c_{XZA} & -c_{XZC} \\ 0 & \sigma_X^2 & c_{XYC} & c_{XYA} & c_{XZC} & c_{XZA} \\ \hline c_{XYA} & c_{XYC} & \sigma_Y^2 & 0 & c_{YZA} & -c_{YZC} \\ -c_{XYC} & c_{XYA} & 0 & \sigma_Y^2 & c_{YZC} & c_{YZA} \\ \hline c_{XZA} & c_{XZC} & c_{YZA} & c_{YZC} & \sigma_Z^2 & 0 \\ -c_{XZC} & c_{XZA} & -c_{YZC} & c_{YZA} & 0 & \sigma_Z^2 \end{array}\right] \tag{7.3}$$

where the subscript "A" indicates "auto" and "C" indicates "cross" with respect to the real and imaginary parts. The covariance matrix $\tilde{\mathbf{C}}$ for the complex variables $\tilde{X}_i$, $\tilde{Y}_i$ and $\tilde{Z}_i$ is

$$\tilde{\mathbf{C}} = 2 \left[\begin{array}{ccc} \sigma_X^2 & c_{XYA} + j\, c_{XYC} & c_{XZA} + j\, c_{XZC} \\ c_{XYA} - j\, c_{XYC} & \sigma_Y^2 & c_{YZA} + j\, c_{YZC} \\ c_{XZA} - j\, c_{XZC} & c_{YZA} - j\, c_{YZC} & \sigma_Z^2 \end{array}\right] \tag{7.4}$$

(see Section 4.1.4 on page 26 for details on these types of covariance matrices). The inverse of $\mathbf{C}$ can be written

$$\mathbf{C}^{-1} = \left[\begin{array}{cc|cc|cc} a_1 & 0 & a_2 & a_3 & c_2 & c_3 \\ 0 & a_1 & -a_3 & a_2 & -c_3 & c_2 \\ \hline a_2 & -a_3 & b_1 & 0 & b_2 & b_3 \\ a_3 & a_2 & 0 & b_1 & -b_3 & b_2 \\ \hline c_2 & -c_3 & b_2 & -b_3 & c_1 & 0 \\ c_3 & c_2 & b_3 & b_2 & 0 & c_1 \end{array}\right], \tag{7.5}$$

where

$$
\begin{aligned}
a_1 &= (c_{YZA}^2 + c_{YZC}^2 - \sigma_Y^2\sigma_Z^2)/\left|\mathbf{C}\right|^{1/2} \\
a_2 &= (\sigma_Z^2 c_{XYA} - c_{XZA}c_{YZA} - c_{XZC}c_{YZC})/\left|\mathbf{C}\right|^{1/2} \\
a_3 &= (c_{XZC}c_{YZA} - c_{XZA}c_{YZC} - \sigma_Z^2 c_{XYC})/\left|\mathbf{C}\right|^{1/2} \\
b_1 &= (c_{XZA}^2 + c_{XZC}^2 - \sigma_X^2\sigma_Z^2)/\left|\mathbf{C}\right|^{1/2} \\
b_2 &= (\sigma_X^2 c_{YZA} - c_{XYA}c_{XZA} - c_{XYC}c_{XZC})/\left|\mathbf{C}\right|^{1/2} \\
b_3 &= (c_{XYA}c_{XZC} - c_{XYC}c_{XZA} - \sigma_X^2 c_{YZC})/\left|\mathbf{C}\right|^{1/2} \\
c_1 &= (c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)/\left|\mathbf{C}\right|^{1/2} \\
c_2 &= (\sigma_Y^2 c_{XZA} - c_{XYA}c_{YZA} + c_{XYC}c_{YZC})/\left|\mathbf{C}\right|^{1/2} \\
c_3 &= (c_{XYA}c_{YZC} + c_{XYC}c_{YZA} - \sigma_Y^2 c_{XZC})/\left|\mathbf{C}\right|^{1/2}.
\end{aligned}
\tag{7.6}
$$

The determinant of $\mathbf{C}$, $|\mathbf{C}|$, is given by

$$
\left|\mathbf{C}\right| = \left[\sigma_X^2(c_{YZA}^2 + c_{YZC}^2) + \sigma_Y^2(c_{XZA}^2 + c_{XZC}^2) + \sigma_Z^2(c_{XYA}^2 + c_{XYC}^2) - \sigma_X^2\sigma_Y^2\sigma_Z^2 - 2T\right]^2, \tag{7.7}
$$

where

$$
T = c_{XYA}c_{XZA}c_{YZA} + c_{XYA}c_{XZC}c_{YZC} + c_{XYC}c_{XZC}c_{YZA} - c_{XYC}c_{XZA}c_{YZC}. \tag{7.8}
$$

It is convenient to define

$$
\begin{aligned}
d_1^2 &= a_2^2 + a_3^2 \\
d_2^2 &= b_2^2 + b_3^2 \\
d_3^2 &= c_2^2 + c_3^2 \\
d_4 &= (a_2c_2 + a_3c_3)b_2 + (a_2c_3 - a_3c_2)b_3.
\end{aligned}
\tag{7.9}
$$

The joint PDF for $\mathbf{W}_i$ is a circular zero-mean Gaussian of the form

$$
\mathrm{f}(\mathbf{W}_i) = \left[(2\pi)^6\left|\mathbf{C}\right|\right]^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{W}_i^T\mathbf{C}^{-1}\mathbf{W}_i\right). \tag{7.10}
$$

The following vectors are defined

$$
\begin{aligned}
\mathbf{X} &= [X_{\mathrm{R}1}, X_{\mathrm{I}1}, X_{\mathrm{R}2}, X_{\mathrm{I}2}, \ldots, X_{\mathrm{R}n}, X_{\mathrm{I}n}]^T \\
\mathbf{Y} &= [Y_{\mathrm{R}1}, Y_{\mathrm{I}1}, Y_{\mathrm{R}2}, Y_{\mathrm{I}2}, \ldots, Y_{\mathrm{R}n}, Y_{\mathrm{I}n}]^T \\
\mathbf{Z} &= [Z_{\mathrm{R}1}, Z_{\mathrm{I}1}, Z_{\mathrm{R}2}, Z_{\mathrm{I}2}, \ldots, Z_{\mathrm{R}n}, Z_{\mathrm{I}n}]^T.
\end{aligned}
\tag{7.11}
$$

For an n-look system, the intensities I_{X}, I_{Y} and I_{Z} are given by

$$\begin{aligned} I_{\mathrm{X}} = A_{\mathrm{X}}^2 &= \frac{\sum_{i=1}^{n}\left(X_{\mathrm{R}i}^2 + X_{\mathrm{I}i}^2\right)}{n} = \frac{|\mathbf{X}|^2}{n} = \frac{\mathcal{A}_{\mathrm{X}}^2}{n} \\ I_{\mathrm{Y}} = A_{\mathrm{Y}}^2 &= \frac{\sum_{i=1}^{n}\left(Y_{\mathrm{R}i}^2 + Y_{\mathrm{I}i}^2\right)}{n} = \frac{|\mathbf{Y}|^2}{n} = \frac{\mathcal{A}_{\mathrm{Y}}^2}{n} \\ I_{\mathrm{Z}} = A_{\mathrm{Z}}^2 &= \frac{\sum_{i=1}^{n}\left(Z_{\mathrm{R}i}^2 + Z_{\mathrm{I}i}^2\right)}{n} = \frac{|\mathbf{Z}|^2}{n} = \frac{\mathcal{A}_{\mathrm{Z}}^2}{n}, \end{aligned} \tag{7.12}$$

where A_{X}, A_{Y} and A_{Z} are the amplitudes and $|\cdot|$ denotes the modulus. $\mathcal{A} = nA$ is the summed, rather than the averaged, amplitude and is introduced for convenience.

The joint density of the summed amplitudes can be computed as the marginal density [Miller et al. 1958]

$$\mathrm{f}(\mathcal{A}_{\mathrm{X}}, \mathcal{A}_{\mathrm{Y}}, \mathcal{A}_{\mathrm{Z}}) = \left[(2\pi)^6 |\mathbf{C}|\right]^{-n/2} \int_{|\mathbf{Z}|=\mathcal{A}_{\mathrm{Z}}} \int_{|\mathbf{Y}|=\mathcal{A}_{\mathrm{Y}}} \int_{|\mathbf{X}|=\mathcal{A}_{\mathrm{X}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \mathbf{W}_i^T \mathbf{C}^{-1} \mathbf{W}_i\right) d\mathbf{X} d\mathbf{Y} d\mathbf{Z}, \tag{7.13}$$

where $\int_{|\mathbf{Z}|=\mathcal{A}}$ denotes integration over the surface of a $2n$-dimensional sphere of radius $\mathcal{A}$.

In the next section the inner integral of Equation 7.13 is solved.

7.1.1 The X integral in Equation 7.13

Using the definition of $\mathbf{W}_i$ in Equation 7.2 and Equation 7.5 on page 53 shows that

$$\begin{aligned} \mathbf{W}_i^T \mathbf{C}^{-1} \mathbf{W}_i &= a_1(X_{Ri}^2 + X_{Ii}^2) + b_1(Y_{Ri}^2 + Y_{Ii}^2) + c_1(Z_{Ri}^2 + Z_{Ii}^2) \\ &\quad + X_{Ri}(a_2 Y_{Ri} + a_3 Y_{Ii} + c_2 Z_{Ri} + c_3 Z_{Ii} + a_2 Y_{Ri} + a_3 Y_{Ii} + c_2 Z_{Ri} + c_3 Z_{Ii}) \\ &\quad + X_{Ii}(-a_3 Y_{Ri} + a_2 Y_{Ii} - c_3 Z_{Ri} + c_2 Z_{Ii} - a_3 Y_{Ri} + a_2 Y_{Ii} - c_3 Z_{Ri} + c_2 Z_{Ii}) \\ &\quad + 2b_2 Y_{Ii} Z_{Ii} + 2b_2 Y_{Ri} Z_{Ri} + 2b_3 Y_{Ri} Z_{Ii} - 2b_3 Y_{Ii} Z_{Ri} \\ &= a_1(X_{Ri}^2 + X_{Ii}^2) + b_1(Y_{Ri}^2 + Y_{Ii}^2) + c_1(Z_{Ri}^2 + Z_{Ii}^2) \\ &\quad + 2X_{Ri}(a_2 Y_{Ri} + a_3 Y_{Ii} + c_2 Z_{Ri} + c_3 Z_{Ii}) \\ &\quad + 2X_{Ii}(a_2 Y_{Ii} - a_3 Y_{Ri} + c_2 Z_{Ii} - c_3 Z_{Ri}) \\ &\quad + 2Y_{Ri}(b_2 Z_{Ri} + b_3 Z_{Ii}) + 2Y_{Ii}(b_2 Z_{Ii} - b_3 Z_{Ri}). \end{aligned} \tag{7.14}$$

The argument of the exponential function in Equation 7.13 can now be expressed in the form

$$-\frac{1}{2}\sum_{i=1}^{n} \mathbf{W}_i^T \mathbf{C}^{-1} \mathbf{W}_i = -\frac{1}{2}\left(a_1 \mathcal{A}_{\mathrm{X}}^2 + b_1 \mathcal{A}_{\mathrm{Y}}^2 + c_1 \mathcal{A}_{\mathrm{Z}}^2 - 2\mathbf{V}\cdot\mathbf{X} - 2\mathbf{T}\cdot\mathbf{Y}\right), \tag{7.15}$$

where

$$\begin{aligned}
\mathbf{V} &= -a_2\mathbf{Y} - a_3\mathbf{Y}^+ - c_2\mathbf{Z} - c_3\mathbf{Z}^+ \\
\mathbf{T} &= -b_2\mathbf{Z} - b_3\mathbf{Z}^+ \\
\mathbf{Y}^+ &= [Y_{I1}, -Y_{R1}, \cdots, Y_{In}, -Y_{Rn}]^T \\
\mathbf{Z}^+ &= [Z_{I1}, -Z_{R1}, \cdots, Z_{In}, -Z_{Rn}]^T.
\end{aligned} \tag{7.16}$$

Since $\mathbf{T}$ is independent of $\mathbf{X}$, Equation 7.13 can be written as

$$\begin{aligned}
\mathrm{f}(\mathcal{A}_\mathrm{X}, \mathcal{A}_\mathrm{Y}, \mathcal{A}_\mathrm{Z}) &= \left[(2\pi)^6 |\mathbf{C}|\right]^{-n/2} \exp\left(-\frac{1}{2}\left(a_1\mathcal{A}_\mathrm{X}^2 + b_1\mathcal{A}_\mathrm{Y}^2 + c_1\mathcal{A}_\mathrm{Z}^2\right)\right) \\
&\quad \times \int_{|\mathbf{Z}|=\mathcal{A}_\mathrm{Z}} \int_{|\mathbf{Y}|=\mathcal{A}_\mathrm{Y}} \exp(\mathbf{T}\cdot\mathbf{Y}) \int_{|\mathbf{X}|=\mathcal{A}_\mathrm{X}} \exp(\mathbf{V}\cdot\mathbf{X})\, d\mathbf{X} d\mathbf{Y} d\mathbf{Z}.
\end{aligned} \tag{7.17}$$

Substituting the result (c.f. Miller et al. [1958])

$$\int_{|\mathbf{X}|=\mathcal{A}_\mathrm{X}} \exp(\mathbf{V}\cdot\mathbf{X})\, d\mathbf{X} = (2\pi\mathcal{A}_\mathrm{X})^n |\mathbf{V}|^{-n+1} \mathrm{I}_{n-1}(|\mathbf{V}|\mathcal{A}_\mathrm{X}) \tag{7.18}$$

into Equation 7.17 gives

$$\begin{aligned}
\mathrm{f}(\mathcal{A}_\mathrm{X}, \mathcal{A}_\mathrm{Y}, \mathcal{A}_\mathrm{Z}) &= \left[(2\pi)^6 |\mathbf{C}|\right]^{-n/2} (2\pi\mathcal{A}_\mathrm{X})^n \exp\left(-\frac{1}{2}\left(a_1\mathcal{A}_\mathrm{X}^2 + b_1\mathcal{A}_\mathrm{Y}^2 + c_1\mathcal{A}_\mathrm{Z}^2\right)\right) \\
&\quad \times \int_{|\mathbf{Z}|=\mathcal{A}_\mathrm{Z}} \int_{|\mathbf{Y}|=\mathcal{A}_\mathrm{Y}} \exp\left((-b_2\mathbf{Z} - b_3\mathbf{Z}^+)\cdot\mathbf{Y}\right) |\mathbf{V}|^{-n+1} \mathrm{I}_{n-1}(|\mathbf{V}|\mathcal{A}_\mathrm{X})\, d\mathbf{Y} d\mathbf{Z},
\end{aligned} \tag{7.19}$$

where $\mathrm{I}_n(\cdot)$ denotes the modified Bessel function of order n.

In order to solve the next integral, the $\mathbf{Y}$ integral from Equation 7.19, two different cases have to be considered. The first case is for $n > 1$ for which results from Miller et al. [1958] can be used; in the second case, for $n = 1$, Miller's solution does not apply and this case is solved in Section 7.1.3.

7.1.2 The Y integral in Equation 7.19 for $n > 1$

For the case $n > 1$, the integral over $\mathbf{Y}$ in Equation 7.19 can be evaluated as follows. The $\mathbf{Y}$ integral in Equation 7.19 is

$$I_1 = \int_{|\mathbf{Y}|=\mathcal{A}_\mathrm{Y}} \exp\left((-b_2\mathbf{Z} - b_3\mathbf{Z}^+)\cdot\mathbf{Y}\right) |\mathbf{V}|^{-n+1} \mathrm{I}_{n-1}(|\mathbf{V}|\mathcal{A}_\mathrm{X})\, d\mathbf{Y}. \tag{7.20}$$

This integral can be evaluated by making use of a result of Miller et al. [1958] for the integral

$$I_2 = \int_{|\mathbf{Y}|=\mathcal{A}_Y} \exp\left((E_{12}\mathbf{A} + E_{22}\mathbf{B}) \cdot \mathbf{Y}\right) |\mathbf{V}|^{-n+1} \mathrm{I}_{n-1}(|\mathbf{V}|\,\mathcal{A}_X) d\mathbf{Y}, \tag{7.21}$$

where $n > 1$ and the vectors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Y}$ and $\mathbf{V}$ and the constants E_{12} and E_{22} are subject to the restriction

$$\mathbf{V} = E_{11}\mathbf{A} + E_{12}\mathbf{B} - E_{12}\mathbf{Y}, \tag{7.22}$$

where E_{11} is an arbitrary parameter, and $\mathcal{A}_X > 0$. They showed that

$$I_2 = \frac{(2\pi)^n 2^{n-1}\Gamma(n-1)\mathcal{A}_Y}{(E_{12}\omega_1\omega_2)^{n-1}\mathcal{A}_X^{n-1}} \sum_{k=n-1}^{\infty} k\ (-1)^{k-n+1} C_{k-n+1}^{n-1}(\cos\gamma)\mathrm{I}_k(E_{12}\mathcal{A}_X\mathcal{A}_Y)\mathrm{I}_k(\omega_1\mathcal{A}_X)\mathrm{I}_k(\omega_2\mathcal{A}_Y), \tag{7.23}$$

where $C_{k-n+1}^{n-1}(\cos\gamma)$ is the Gegenbauer function (or ultraspherical polynomial see Appendix A on page 143),

$$\omega_1 = |E_{11}\mathbf{A} + E_{12}\mathbf{B}| \tag{7.24}$$

$$\omega_2 = |E_{12}\mathbf{A} + E_{22}\mathbf{B}| \tag{7.25}$$

and γ is the angle between the vectors $E_{11}\mathbf{A} + E_{12}\mathbf{B}$ and $E_{12}\mathbf{A} + E_{22}\mathbf{B}$ and therefore

$$\cos\gamma = \frac{(E_{11}\mathbf{A} + E_{12}\mathbf{B}) \cdot (E_{12}\mathbf{A} + E_{22}\mathbf{B})}{|E_{11}\mathbf{A} + E_{12}\mathbf{B}|\,|E_{12}\mathbf{A} + E_{22}\mathbf{B}|}. \tag{7.26}$$

Note that $\mathbf{A}$ and $\mathbf{B}$ do not appear in Equation 7.23.

To show that the result Equation 7.23 can be used in this case the following steps are taken. Matching Equation 7.20 and Equation 7.21, and using Equation 7.16 gives

$$E_{12}\mathbf{A} + E_{22}\mathbf{B} = -b_2\mathbf{Z} - b_3\mathbf{Z}^+ = \mathbf{T} \tag{7.27}$$

and using Equation 7.16, Equation 7.22 becomes

$$\mathbf{V} = E_{11}\mathbf{A} + E_{12}\mathbf{B} - E_{12}\mathbf{Y} = -a_2\mathbf{Y} - a_3\mathbf{Y}^+ - c_2\mathbf{Z} - c_3\mathbf{Z}^+. \tag{7.28}$$

Equations 7.27 and 7.28 provide two linear equations for $\mathbf{A}$ and $\mathbf{B}$ so that it is possible to solve for $\mathbf{A}$ and $\mathbf{B}$. Therefore, Equation 7.21 has the same structure as Equation 7.20, and Equation 7.23 can be used to evaluate Equation 7.20. Now the parameters in Equation 7.23 are evaluated.

Using Equation 7.27, substituting for $\mathbf{Z}^+$ from Equation 7.16, and using the definitions of d_2

and $\mathcal{A}_Z$ from Equations 7.9 and 7.12, respectively, shows that

$$\begin{aligned}
\omega_2^2 &= |E_{12}\mathbf{A} + E_{22}\mathbf{B}|^2 = |T|^2 \\
&= \left|-b_2\mathbf{Z} - b_3\mathbf{Z}^+\right|^2 \\
&= \sum_{i=1}^{n}\left[(-b_2 Z_{Ri} - b_3 Z_{Ii})^2 + (-b_2 Z_{Ii} + b_3 Z_{Ri})^2\right] \\
&= \sum_{i=1}^{n}\left[\left(b_2^2 + b_3^2\right) Z_{Ri} + \left(b_2^2 + b_3^2\right) Z_{Ii}\right] \\
&= d_2^2\mathcal{A}_Z^2. \qquad (7.29)
\end{aligned}$$

From Equation 7.22

$$|\mathbf{V}|^2 = E_{12}^2\mathcal{A}_Y^2 + \omega_1^2 - 2E_{12}(E_{11}\mathbf{A} + E_{12}\mathbf{B})\cdot\mathbf{Y}. \qquad (7.30)$$

Using the right hand side of Equation 7.28, and substituting from 7.12 and 7.16, allows $|\mathbf{V}|^2$ to be also written in the form

$$\begin{aligned}
|\mathbf{V}|^2 &= \left|-a_2\mathbf{Y} - a_3\mathbf{Y}^+ - c_2\mathbf{Z} - c_3\mathbf{Z}^+\right|^2 \\
&= \sum_{i=1}^{n}\left[(-a_2 Y_{Ri} - a_3 Y_{Ii} - c_2 Z_{Ri} - c_3 Z_{Ii})^2 + (-a_2 Y_{Ii} + a_3 Y_{Ri} - c_2 Z_{Ii} + c_3 Z_{Ri})^2\right] \\
&= d_1^2\mathcal{A}_Y^2 + d_3^2\mathcal{A}_Z^2 + 2\big(a_2c_2 Y_{Ri}Z_{Ri} + a_2c_3 Y_{Ri}Z_{Ii} + a_3c_2 Y_{Ii}Z_{Ri} + a_3c_3 Y_{Ii}Z_{Ii} \\
&\quad + a_2c_2 Y_{Ii}Z_{Ii} - a_2c_3 Y_{Ii}Z_{Ri} - a_3c_2 Y_{Ri}Z_{Ii} + a_3c_3 Y_{Ri}Z_{Ri}\big) \\
&= d_1^2\mathcal{A}_Y^2 + d_3^2\mathcal{A}_Z^2 + 2\left((a_2c_2 + a_3c_3)\mathbf{Z} + (a_2c_3 - a_3c_2)\mathbf{Z}^+\right)\cdot\mathbf{Y} \qquad (7.31)
\end{aligned}$$

or

$$|\mathbf{V}|^2 = d_1^2\mathcal{A}_Y^2 + d_3^2\mathcal{A}_Z^2 + \mathbf{U}\cdot\mathbf{Y}, \qquad (7.32)$$

where

$$\mathbf{U} = 2\left((a_2c_2 + a_3c_3)\mathbf{Z} + (a_2c_3 - a_3c_2)\mathbf{Z}^+\right). \qquad (7.33)$$

It follows from 7.30 and 7.32 that

$$\begin{aligned}
E_{12} &= d_1 \\
\omega_1 &= |E_{11}\mathbf{A} + E_{12}\mathbf{B}| = d_3\mathcal{A}_Z.
\end{aligned} \qquad (7.34)$$

From the definition of γ, $\cos\gamma$ is given by

$$\cos\gamma = \frac{(E_{11}\mathbf{A} + E_{12}\mathbf{B})\cdot(E_{12}\mathbf{A} + E_{22}\mathbf{B})}{|E_{11}\mathbf{A} + E_{12}\mathbf{B}|\,|E_{12}\mathbf{A} + E_{22}\mathbf{B}|}. \qquad (7.35)$$

Comparison of Equations 7.30 and 7.32 shows that

$$\mathbf{U} = -2E_{12}(E_{11}\mathbf{A} + E_{12}\mathbf{B}), \qquad (7.36)$$

and substituting into Equation 7.35 and using Equation 7.27 gives

$$\cos\gamma = \frac{\frac{\mathbf{U}}{-2E_{12}} \cdot \mathbf{T}}{\omega_1\omega_2} = \frac{\mathbf{U}\cdot\mathbf{T}}{-2d_1d_2d_3\mathcal{A}_{\mathrm{Z}}^2}. \tag{7.37}$$

Using 7.27 and 7.33 gives

$$\begin{aligned}
\mathbf{U}\cdot\mathbf{T} &= 2\left((a_2c_2+a_3c_3)\mathbf{Z}+(a_2c_3-a_3c_2)\mathbf{Z}^+\right)\cdot\left(-b_2\mathbf{Z}-b_3\mathbf{Z}^+\right)\\
&= -2\left((a_2c_2+a_3c_3)b_2\mathbf{Z}\cdot\mathbf{Z}+(a_2c_2+a_3c_3)b_3\mathbf{Z}\cdot\mathbf{Z}^+\right.\\
&\quad \left.+(a_2c_3-a_3c_2)b_2\mathbf{Z}^+\cdot\mathbf{Z}+(a_2c_3-a_3c_2)b_3\mathbf{Z}^+\cdot\mathbf{Z}^+\right)\\
&= -2\big((a_2c_2+a_3c_3)b_2+(a_2c_3-a_3c_2)b_3\big)\mathcal{A}_{\mathrm{Z}}^2\\
&= -2d_4\mathcal{A}_{\mathrm{Z}}^2.
\end{aligned} \tag{7.38}$$

Note that $\mathbf{Z}\cdot\mathbf{Z}^+ = \mathbf{Z}^+\cdot\mathbf{Z} = 0$ and $\mathbf{Z}\cdot\mathbf{Z} = \mathbf{Z}^+\cdot\mathbf{Z}^+ = \mathcal{A}_{\mathrm{Z}}^2$. With this Equation 7.37 becomes

$$\cos\gamma = \frac{-2d_4\mathcal{A}_{\mathrm{Z}}^2}{-2d_1d_2d_3\mathcal{A}_{\mathrm{Z}}^2} = \frac{d_4}{d_1d_2d_3}. \tag{7.39}$$

Substituting the above expressions for E_{12}, ω_1 and ω_2 into Equation 7.23 allows Equation 7.20 to be evaluated as

$$\begin{aligned}
&I_1 =\\
&\frac{(2\pi)^n\,2^{n-1}\Gamma(n-1)\mathcal{A}_{\mathrm{Y}}}{(d_1d_2d_3)^{n-1}\mathcal{A}_{\mathrm{X}}^{n-1}\mathcal{A}_{\mathrm{Z}}^{2n-2}}\sum_{k=n-1}^{\infty}k\,(-1)^{k-n+1}\,C_{k-n+1}^{n-1}(\cos\gamma)\mathrm{I}_k\left(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}}\right)\mathrm{I}_k\left(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\right)\mathrm{I}_k\left(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}}\right),
\end{aligned} \tag{7.40}$$

and substituting into Equation 7.19 gives

$$\begin{aligned}
\mathrm{f}(\mathcal{A}_{\mathrm{X}},\mathcal{A}_{\mathrm{Y}},\mathcal{A}_{\mathrm{Z}}) &= \left[(2\pi)^6\,|\mathbf{C}|\right]^{-n/2}(2\pi\mathcal{A}_{\mathrm{X}})^n\exp\left(-\frac{1}{2}\left(a_1\mathcal{A}_{\mathrm{X}}^2+b_1\mathcal{A}_{\mathrm{Y}}^2+c_1\mathcal{A}_{\mathrm{Z}}^2\right)\right)\\
&\quad\times\int_{|\mathbf{Z}|=\mathcal{A}_{\mathrm{Z}}}\frac{\mathcal{A}_{\mathrm{Y}}\,(2\pi)^n\,2^{n-1}\Gamma(n-1)}{(d_1d_2d_3)^{n-1}\mathcal{A}_{\mathrm{X}}^{n-1}\mathcal{A}_{\mathrm{Z}}^{2n-2}}\sum_{k=n-1}^{\infty}k\,(-1)^{k-n+1}\,C_{k-n+1}^{n-1}(\cos\gamma)\\
&\quad\times\mathrm{I}_k\left(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}}\right)\mathrm{I}_k\left(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\right)\mathrm{I}_k\left(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}}\right)d\mathbf{Z},
\end{aligned} \tag{7.41}$$

where $C_{k-n+1}^{n-1}(\cos\gamma)$ is the Gegenbauer function (or ultraspherical polynomial see Appendix A on page 143).

The integrand in Equation 7.41 is independent of $\mathbf{Z}$, and

$$\int_{|\mathbf{Z}|=\mathcal{A}} 1\,d\mathbf{Z} = \frac{2\pi^n}{\Gamma(n)}\mathcal{A}^{2n-1}, \tag{7.42}$$

where $\Gamma(\cdot)$ is the Gamma function. Therefore Equation 7.41 becomes

$$\begin{aligned} \mathrm{f}(\mathcal{A}_\mathrm{X},\mathcal{A}_\mathrm{Y},\mathcal{A}_\mathrm{Z}) &= \frac{\exp\left(-\frac{1}{2}\left(a_1\mathcal{A}_\mathrm{X}^2+b_1\mathcal{A}_\mathrm{Y}^2+c_1\mathcal{A}_\mathrm{Z}^2\right)\right)\mathcal{A}_\mathrm{X}\mathcal{A}_\mathrm{Y}\mathcal{A}_\mathrm{Z}}{(n-1)\,|\mathbf{C}|^{n/2}\,(d_1d_2d_3)^{n-1}} \\ &\quad\times\sum_{k=n-1}^{\infty} k\ (-1)^{k-n+1}\,C_{k-n+1}^{n-1}(\cos\gamma)\mathrm{I}_k\left(d_1\mathcal{A}_\mathrm{X}\mathcal{A}_\mathrm{Y}\right)\mathrm{I}_k\left(d_2\mathcal{A}_\mathrm{Y}\mathcal{A}_\mathrm{Z}\right)\mathrm{I}_k\left(d_3\mathcal{A}_\mathrm{X}\mathcal{A}_\mathrm{Z}\right). \end{aligned} \tag{7.43}$$

The joint density for the intensities is then computed by use of the Jacobian to convert the amplitudes $\mathcal{A}$ to the averaged intensities I ($I=\mathcal{A}^2/n$). In this case the transformation is

$$\mathrm{f}(I_X,I_Y,I_Z)=\frac{1}{8}\frac{n^{3/2}}{\sqrt{I_XI_YI_Z}}\mathrm{f}(\mathcal{A}_\mathrm{X},\mathcal{A}_\mathrm{Y},\mathcal{A}_\mathrm{Z})\Big|_{\mathcal{A}_\mathrm{X}=\sqrt{nI_X},\ \mathcal{A}_\mathrm{Y}=\sqrt{nI_Y},\ \mathcal{A}_\mathrm{Z}=\sqrt{nI_Z}}. \tag{7.44}$$

The density for the intensity values is therefore

$$\begin{aligned} \mathrm{f}(I_\mathrm{X},I_\mathrm{Y},I_\mathrm{Z}) &= \frac{n^3\exp\left(-\frac{n}{2}\left(a_1I_\mathrm{X}+b_1I_\mathrm{Y}+c_1I_\mathrm{Z}\right)\right)}{8(n-1)\,|\mathbf{C}|^{n/2}\,(d_1d_2d_3)^{n-1}} \\ &\quad\times\sum_{k=n-1}^{\infty} k(-1)^{k-n+1}C_{k-n+1}^{n-1}(\cos\gamma)\mathrm{I}_k\left(d_1n\sqrt{I_\mathrm{X}I_\mathrm{Y}}\right)\mathrm{I}_k\left(d_2n\sqrt{I_\mathrm{Y}I_\mathrm{Z}}\right)\mathrm{I}_k\left(d_3n\sqrt{I_\mathrm{X}I_\mathrm{Z}}\right). \end{aligned} \tag{7.45}$$

7.1.3 The Y integral in Equation 7.19 for $n = 1$

For the case $n = 1$, $\mathbf{Y}$ has only two components and the integral over $\mathbf{Y}$ in Equation 7.19 on page 56 is over a circle rather than over a $2n$-dimensional sphere. The derivation in Section 7.1.2 does not then apply, and this case is evaluated here. The $\mathbf{Y}$ integral in Equation 7.19 for the case $n = 1$ is

$$I=\int_{|\mathbf{Y}|=\mathcal{A}_\mathrm{Y}}\exp\left(\left(-b_2\mathbf{Z}-b_3\mathbf{Z}^+\right)\cdot\mathbf{Y}\right)\mathrm{I}_0\left(|\mathbf{V}|\,\mathcal{A}_\mathrm{X}\right)d\mathbf{Y}. \tag{7.46}$$

Using Equations 7.27 and 7.32 on page 58 allows the integral to be written in the form

$$I=\int_{|\mathbf{Y}|=\mathcal{A}_\mathrm{Y}}\exp\left(\mathbf{T}\cdot\mathbf{Y}\right)\mathrm{I}_0\left(\sqrt{d_1^2\mathcal{A}_\mathrm{X}^2\mathcal{A}_\mathrm{Y}^2+d_3^2\mathcal{A}_\mathrm{X}^2\mathcal{A}_\mathrm{Z}^2+\mathcal{A}_\mathrm{X}^2\mathbf{U}\cdot\mathbf{Y}}\right)d\mathbf{Y}. \tag{7.47}$$

The integral is in the plane and the following angles θ, ϕ and ψ are defined in this plane by

$$\begin{aligned} \mathbf{Y} &= [Y_{R1},\ Y_{I1}]^T=[\mathcal{A}_\mathrm{Y}\cos\theta\ ,\ \mathcal{A}_\mathrm{Y}\sin\theta]^T \text{ with } |\mathbf{Y}|=\mathcal{A}_\mathrm{Y} \\ \mathbf{U} &= [|\mathbf{U}|\cos\phi\ ,\ |\mathbf{U}|\sin\phi]^T \\ \mathbf{T} &= [|\mathbf{T}|\cos\psi\ ,\ |\mathbf{T}|\sin\psi]^T. \end{aligned} \tag{7.48}$$

Equation 7.29 on page 58 gives

$$|\mathbf{T}| = \omega_2 = d_2 \mathcal{A}_{\mathrm{Z}}, \tag{7.49}$$

and using the definitions of $\mathbf{Z}$ (Equation 7.11 on page 54) and $\mathbf{Z}^+$ (Equation 7.16 on page 56) and the newly defined substitutions $d_5 = a_2c_2 + a_3c_3$ and $d_6 = a_2c_3 - a_3c_2$ shows that

$$\begin{aligned} |\mathbf{U}|^2 &= \left|2\left(d_5\mathbf{Z} + d_6\mathbf{Z}^+\right)\right|^2 \\ &= \sum_{i=1}^{n}\left[\left(2d_5Z_{Ri} + 2d_6Z_{Ii}\right)^2 + \left(2d_5Z_{Ii} - 2d_6Z_{Ri}\right)^2\right] \\ &= \sum_{i=1}^{n}\left[4\left(d_5^2 + d_6^2\right)Z_{Ri}^2 + 4\left(d_5^2 + d_6^2\right)Z_{Ii}^2\right] \\ &= 4\left(d_5^2 + d_6^2\right)\mathcal{A}_{\mathrm{Z}}^2 \\ &= 4d_1^2d_3^2\mathcal{A}_{\mathrm{Z}}^2 \end{aligned} \tag{7.50}$$

(note that $d_5^2 + d_6^2 = d_1^2d_3^2$). Therefore,

$$\begin{aligned} \mathbf{U}\cdot\mathbf{Y} &= -2d_1d_3\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\cos(\theta - \phi) \\ \mathbf{T}\cdot\mathbf{Y} &= d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\cos(\theta - \psi). \end{aligned} \tag{7.51}$$

Converting Graf's addition theorem (Abramowitz and Stegun [1972], page 363) for the Bessel function $\mathrm{J}_0(\cdot)$

$$\mathrm{J}_0\left(\sqrt{u^2 + v^2 - 2uv\cos\alpha}\right) = \sum_{k=-\infty}^{\infty}\mathrm{J}_k(u)\mathrm{J}_k(v)\cos k\alpha \tag{7.52}$$

into a version with modified Bessel functions

$$\mathrm{I}_0\left(\sqrt{u^2 + v^2 - 2uv\cos\alpha}\right) = \sum_{k=-\infty}^{\infty}(-1)^k\mathrm{I}_k(u)\mathrm{I}_k(v)\cos k\alpha. \tag{7.53}$$

Making the substitutions

$$\begin{aligned} u &= d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}} \\ v &= d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}} \\ \alpha &= \theta - \phi \end{aligned} \tag{7.54}$$

in Equation 7.53, using Equation 7.51, and interchanging the order of summation and integration in Equation 7.47, shows that

$$I = \mathcal{A}_{\mathrm{Y}}\sum_{k=-\infty}^{\infty}(-1)^k\mathrm{I}_k(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}})\mathrm{I}_k(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}})\int_0^{2\pi}\exp\left(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\cos(\theta - \psi)\right)\cos(k(\theta - \phi))d\theta. \tag{7.55}$$

The integral is evaluated using the integral form of the modified Bessel function (Abramowitz and Stegun [1972], page 360) giving

$$I = 2\pi\mathcal{A}_{\mathrm{Y}}\Big[\mathrm{I}_0(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}})\mathrm{I}_0(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}})\mathrm{I}_0(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}}) + 2\sum_{k=1}^{\infty}\cos(k(\psi-\phi))\mathrm{I}_k(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}})\mathrm{I}_k(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}})\mathrm{I}_k(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}})\Big]. \quad (7.56)$$

Noting that $\psi-\phi$ is the angle between $\mathbf{U}$ and $\mathbf{T}$, $\psi-\phi=\gamma$, and using Equation 7.56 in Equation 7.19 leads to

$$\begin{aligned}\mathrm{f}(\mathcal{A}_{\mathrm{X}},\mathcal{A}_{\mathrm{Y}},\mathcal{A}_{\mathrm{Z}}) &= \left[(2\pi)^6|\mathbf{C}|\right]^{-n/2}(2\pi\mathcal{A}_{\mathrm{X}})^n\exp\left(-\frac{1}{2}\left(a_1\mathcal{A}_{\mathrm{X}}^2+b_1\mathcal{A}_{\mathrm{Y}}^2+c_1\mathcal{A}_{\mathrm{Z}}^2\right)\right)\\ &\quad\times\int_{|\mathbf{Z}|=\mathcal{A}_{\mathrm{Z}}}2\pi\mathcal{A}_{\mathrm{Y}}\Big[\mathrm{I}_0(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}})\mathrm{I}_0(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}})\mathrm{I}_0(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}})\\ &\quad+2\sum_{k=1}^{\infty}\cos(k\gamma)\mathrm{I}_k(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}})\mathrm{I}_k(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}})\mathrm{I}_k(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}})\Big]d\mathbf{Z}. \quad (7.57)\end{aligned}$$

The integral in Equation 7.57 is independent of $\mathbf{Z}$ and is calculated as in the evaluation of Equation 7.42. This leads the to joint density

$$\begin{aligned}\mathrm{f}(\mathcal{A}_{\mathrm{X}},\mathcal{A}_{\mathrm{Y}},\mathcal{A}_{\mathrm{Z}}) &= |\mathbf{C}|^{-\frac{1}{2}}\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}}\mathcal{A}_Z\exp\left(-\frac{1}{2}\left[a_1\mathcal{A}_{\mathrm{X}}^2+b_1\mathcal{A}_{\mathrm{Y}}^2+c_1\mathcal{A}_{\mathrm{Z}}^2\right]\right)\\ &\quad\times\Big[\mathrm{I}_0\left(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}}\right)\mathrm{I}_0\left(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\right)\mathrm{I}_0\left(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}}\right)\\ &\quad+2\sum_{k=1}^{\infty}\cos(k\gamma)\mathrm{I}_k\left(d_1\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Y}}\right)\mathrm{I}_k\left(d_2\mathcal{A}_{\mathrm{Y}}\mathcal{A}_{\mathrm{Z}}\right)\mathrm{I}_k\left(d_3\mathcal{A}_{\mathrm{X}}\mathcal{A}_{\mathrm{Z}}\right)\Big]. \quad (7.58)\end{aligned}$$

The joint density of the averaged intensities for the case $n=1$ is evaluated using the transformation in Equation 7.44 on page 60 and is

$$\begin{aligned}\mathrm{f}(I_{\mathrm{X}},I_{\mathrm{Y}},I_{\mathrm{Z}}) &= \frac{\exp\left(-\frac{1}{2}\left(a_1I_{\mathrm{X}}+b_1I_{\mathrm{Y}}+c_1I_{\mathrm{Z}}\right)\right)}{8|\mathbf{C}|^{1/2}}\Big[\mathrm{I}_0\left(d_1\sqrt{I_{\mathrm{X}}I_{\mathrm{Y}}}\right)\mathrm{I}_0\left(d_2\sqrt{I_{\mathrm{Y}}I_{\mathrm{Z}}}\right)\mathrm{I}_0\left(d_3\sqrt{I_{\mathrm{X}}I_{\mathrm{Z}}}\right)\\ &\quad+2\sum_{k=1}^{\infty}\cos(k\gamma)\mathrm{I}_k\left(d_1\sqrt{I_{\mathrm{X}}I_{\mathrm{Y}}}\right)\mathrm{I}_k\left(d_2\sqrt{I_{\mathrm{Y}}I_{\mathrm{Z}}}\right)\mathrm{I}_k\left(d_3\sqrt{I_{\mathrm{X}}I_{\mathrm{Z}}}\right)\Big]. \quad (7.59)\end{aligned}$$

7.2 SPECIAL CASES OF THE JOINT DENSITY

Three special cases of the joint density are discussed here. It is shown that the joint density reduces to the correct forms for these three special cases.

7.2.1 The uncorrelated case

Consider $\tilde{X}_i$, $\tilde{Y}_i$ and $\tilde{Z}_i$ mutually uncorrelated. The covariance matrix $\mathbf{C}$ is then diagonal (all off-diagonal values are zero), as is $\mathbf{C}^{-1}$, and the elements of $\mathbf{C}^{-1}$ (c.f. Equation 7.5 on page 53) become

$$\begin{array}{lll} a_1 = 1/\sigma_X^2 & b_1 = 1/\sigma_Y^2 & c_1 = 1/\sigma_Z^2 \\ a_2 = 0 & b_2 = 0 & c_2 = 0 \\ a_3 = 0 & b_3 = 0 & c_3 = 0. \end{array} \tag{7.60}$$

From 7.60 it follows that

$$d_1 = 0 \qquad d_2 = 0 \qquad d_3 = 0 \qquad d_4 = 0. \tag{7.61}$$

These values cannot be substituted directly into Equation 7.45. However looking at one of the terms in the Equation 7.45, for example

$$\frac{\mathrm{I}_k\left(d_1 n\sqrt{I_X I_Y}\right)}{d_1^{n-1}} \tag{7.62}$$

and using the power series expansion for $\mathrm{I}_k(x)$ (Abramowitz and Stegun [1972], page 375)

$$\frac{\mathrm{I}_k(ax)}{a^{n-1}} = \frac{\left(\frac{1}{2}ax\right)^k}{a^{n-1}} \sum_{l=0}^{\infty} \frac{\left(\frac{1}{4}(ax)^2\right)^l}{l!\Gamma(n+l+1)} \tag{7.63}$$

and substituting $a = d_1$ and $x = n\sqrt{I_X I_Y}$ in Equation 7.63 shows that

$$\lim_{a\to 0} \frac{\mathrm{I}_k(ax)}{a^{n-1}} = \begin{cases} (x/2)^{n-1}/(n-1)! & , \; k = n-1 \\ 0 & , \; k > n-1. \end{cases} \tag{7.64}$$

All terms except for $k = n-1$ therefore vanish in the sum in Equation 7.45. Noting that $C_0^{n-1}(\cos\gamma) = 1$ for all γ and $|\mathbf{C}| = \sigma_X^4\sigma_Y^4\sigma_Z^4$, Equation 7.45 reduces to

$$\mathrm{f}(I_X, I_Y, I_Z) = \prod_{l\in\{X,\,Y,\,Z\}} \frac{n^n I_l^{n-1}}{\sigma_l^{2n}\Gamma(n)\,2^n} \exp\left(-\frac{nI_l}{2\sigma_l^2}\right). \tag{7.65}$$

The density is thus the product of three chi-square densities, as in Equation 4.7 on page 25, as expected.

7.2.2 The two-dimensional case

Removing the correlations between $\tilde{X}_i$ and $\tilde{Z}_i$ and between $\tilde{Y}_i$ and $\tilde{Z}_i$, a two-dimensional correlated distribution (in I_{X} and I_{Y}) and an independent distribution in I_{Z} is obtained. In this case

$$c_{XZA} = c_{XZC} = c_{YZA} = c_{YZC} = 0 \tag{7.66}$$

so that

$$b_2 = b_3 = c_2 = c_3 = 0 \tag{7.67}$$

and therefore

$$d_2 = d_3 = d_4 = 0. \tag{7.68}$$

Using Equation 7.6 gives, in this case,

$$\begin{aligned} a_1 &= -\sigma_Y^2/(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2) \\ b_1 &= -\sigma_X^2/(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2) \\ c_1 &= 1/\sigma_Z^2 \\ d_1^2 &= \left(c_{XYA}^2 + c_{XYC}^2\right)/(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)^2 \\ |\mathbf{C}| &= \sigma_Z^4(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)^2. \end{aligned} \tag{7.69}$$

As in the previous subsection, all terms in Equation 7.45 except for $k = n-1$ vanish, and Equation 7.45 reduces to

$$\mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}}, I_{\mathrm{Z}}) = \frac{n^3 \exp\left(-\frac{n}{2}\left(a_1 I_{\mathrm{X}} + b_1 I_{\mathrm{Y}} + c_1 I_{\mathrm{Z}}\right)\right)}{\left((n-1)!\right)^2 2^{2n+1} d_1^{n-1} |\mathbf{C}|^{n/2}} \left(n^2 I_{\mathrm{Z}} \sqrt{I_{\mathrm{X}} I_{\mathrm{Y}}}\right)^{n-1} \mathrm{I}_{n-1}\left(d_1 n \sqrt{I_{\mathrm{X}} I_{\mathrm{Y}}}\right). \tag{7.70}$$

Equation 7.70 is clearly a product of the form of $\mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}})\mathrm{f}(I_{\mathrm{Z}})$, i.e.

$$\begin{aligned} &\mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}}, I_{\mathrm{Z}}) = \mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}})\, \mathrm{f}(I_{\mathrm{Z}}) = \\ &\left[\frac{n^{n+1}(I_{\mathrm{X}} I_{\mathrm{Y}})^{(n-1)/2} \exp\left(-\frac{n}{2}\left(a_1 I_{\mathrm{X}} + b_1 I_{\mathrm{Y}}\right)\right) \mathrm{I}_{n-1}\left(d_1 n \sqrt{I_{\mathrm{X}} I_{\mathrm{Y}}}\right)}{2^{n+1}\Gamma(n) d_1^{n-1} (c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)^n}\right] \left[\frac{n^n I_{\mathrm{Z}}^{n-1} \exp\left(-\frac{n}{2} c_1 I_{\mathrm{Z}}\right)}{2^n \Gamma(n) \sigma_Z^{2n}}\right]. \end{aligned} \tag{7.71}$$

It follows that

$$\mathrm{f}(I_{\mathrm{Z}}) = \frac{n^n}{2^n \Gamma(n) \sigma_Z^{2n}} \exp\left(-\frac{n I_{\mathrm{Z}}}{2\sigma_Z^2}\right) I_{\mathrm{Z}}^{n-1} \tag{7.72}$$

and

$$\begin{aligned} \mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}}) = &\frac{n^{n+1} \exp\left(\frac{n\left(\sigma_Y^2 I_{\mathrm{X}} + \sigma_X^2 I_{\mathrm{Y}}\right)}{2(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)}\right) (I_{\mathrm{X}} I_{\mathrm{Y}})^{(n-1)/2}}{2^{n+1}\Gamma(n)\left[\sigma_X^2\sigma_Y^2 - (c_{XYA}^2 + c_{XYC}^2)\right]\left(c_{XYA}^2 + c_{XYC}^2\right)^{(n-1)/2}} \\ &\times \mathrm{I}_{n-1}\left(\frac{\sqrt{c_{XYA}^2 + c_{XYC}^2}\sqrt{I_{\mathrm{X}} I_{\mathrm{Y}}}}{\sigma_X^2\sigma_Y^2 - (c_{XYA}^2 + c_{XYC}^2)}\right). \end{aligned} \tag{7.73}$$

The correlation coefficient ρ_c between the complex variables $\tilde{X}_i$ and $\tilde{Y}_i$ is given by

$$|\rho_c|^2 = (c_{XYA}^2 + c_{XYC}^2)/(\sigma_X^2\sigma_Y^2), \tag{7.74}$$

and substituting ρ_c into Equation 7.73 gives

$$\mathrm{f}(I_\mathrm{X}, I_\mathrm{Y}) = \frac{\exp\left(-n\dfrac{I_\mathrm{X}/C_{11} + I_\mathrm{Y}/C_{22}}{1-|\rho_c|^2}\right) n^{n+1}(I_\mathrm{X}I_\mathrm{Y})^{(n-1)/2}}{\left(1-|\rho_c|^2\right)\Gamma(n)|\rho_c|^{n-1}(C_{11}C_{22})^{(n+1)/2}} \mathrm{I}_{n-1}\left(\frac{2n\ |\rho_c|}{1-|\rho_c|^2}\sqrt{\frac{I_\mathrm{X}I_\mathrm{Y}}{C_{11}C_{22}}}\right) \tag{7.75}$$

with $C_{11} = 2\sigma_X^2$ and $C_{22} = 2\sigma_Y^2$. Equation 7.75 corresponds to the two-dimensional case and is equivalent to the result given in Lee et al. [1994b].

7.2.3 The real case

The joint density for the square root of the sum of N squared samples from a *real* three-dimensional zero-mean correlated Gaussian distribution has been derived in Miller et al. [1958]. This case can be made equivalent to ours by replacing N by $n = N/2$, treating the imaginary parts as real variables, and removing the correlations between the real and imaginary parts. This gives

$$c_{XYC} = c_{XZC} = c_{YZC} = 0 \tag{7.76}$$

so that

$$\begin{aligned} a_3 = b_3 = c_3 = 0 \\ \cos\gamma = 1. \end{aligned} \tag{7.77}$$

The Gegenbauer function in this case is $\mathrm{C}_{k-n+1}^{n-1}(1)$ which is the coefficient for the term $\mathcal{K}^{k-n+1}$ in the expansion of $(1-2\mathcal{K}+\mathcal{K}^2)^{-n+1} = (1-\mathcal{K})^{-2n+2}$ (see Appendix A for details of the Gegenbauer function). Which can be expressed in the form of the binomial expansion

$$(1-\mathcal{K})^{-2n+2} = \sum_{l=0}^{-2n+2}\binom{-2n+2}{l}(-\mathcal{K})^l \tag{7.78}$$

and therefore the Gegenbauer function can be simplified to

$$\begin{aligned} \mathrm{C}_{k-n+1}^{n-1}(1) &= \binom{-2n+2}{k-n+1}(-1)^{k-n+1} \\ &= \frac{(-2n+2)(-2n+1)\cdots(-2n+2-k+n-1+1)}{(k-n+1)!}(-1)^{k-n+1} \\ &= \frac{(2n-2)(2n-1)\cdots(n+k-2)}{(k-n+1)!} \\ &= \frac{(n+k-2)!}{(k-n+1)!(2n-3)!} \\ &= \frac{(n+k-2)!}{(k+n-2-2n+3)!(2n-3)!} \\ &= \binom{n+k-2}{2n-3}. \end{aligned} \tag{7.79}$$

With this Equation 7.43 reduces to

$$\begin{aligned} \mathrm{f}(\mathcal{A}_\mathrm{X}, \mathcal{A}_\mathrm{Y}, \mathcal{A}_\mathrm{Z}) &= \frac{\exp\left(-\frac{1}{2}\left(a_1\mathcal{A}_\mathrm{X}^2 + b_1\mathcal{A}_\mathrm{Y}^2 + c_1\mathcal{A}_\mathrm{Z}^2\right)\right)\mathcal{A}_\mathrm{X}\mathcal{A}_\mathrm{Y}\mathcal{A}_\mathrm{Z}}{(n-1)\,|\mathbf{C}|^{n/2}\,(a_2b_2c_2)^{n-1}} \\ &\times \sum_{k=n-1}^{\infty} k(-1)^{k-n+1}\binom{n+k-2}{2n-3}\mathrm{I}_k\left(a_2\mathcal{A}_\mathrm{X}\mathcal{A}_\mathrm{Y}\right)\mathrm{I}_k\left(b_2\mathcal{A}_\mathrm{Y}\mathcal{A}_\mathrm{Z}\right)\mathrm{I}_k\left(c_2\mathcal{A}_\mathrm{X}\mathcal{A}_\mathrm{Z}\right). \end{aligned} \tag{7.80}$$

Making the substitutions

$$\begin{aligned} N &= 2n \\ |\mathbf{M}|^2 &= |\mathbf{C}| \\ M_{12} &= a_2\,|\mathbf{M}| \\ M_{23} &= b_2\,|\mathbf{M}| \\ M_{31} &= c_2\,|\mathbf{M}| \end{aligned} \tag{7.81}$$

shows that Equation 7.80 is identical to Equation 1.10 of Miller et al. [1958].

7.3 THE CHARACTERISTIC FUNCTION

An alternative route to calculating the joint density is via the characteristic function. This approach is pursued here and an alternative form of the joint density is obtained. The results are also used in the next section to derive a method to estimate the parameters of the joint density from intensity data. For more details on the properties of the characteristic function the reader is referred to Papoulis [1991].

The starting point is the result in Krishnamoorthy and Parthasarathy [1951] for the characteristic function for the real case. A straightforward extension of their result to the complex case

shows that the characteristic function is given by

$$\Psi(t_1,t_2,t_3) = \mathrm{E}\big[\exp\big(j(t_1\mathcal{I}_\mathrm{X} + t_2\mathcal{I}_\mathrm{Y} + t_3\mathcal{I}_\mathrm{Z})\big)\big] = \Big|\mathbf{I} - j\,\tilde{\mathbf{C}}\,\mathrm{diag}(t_1,t_2,t_3)\Big|^{-n}, \tag{7.82}$$

where

$$\begin{aligned}\mathcal{I}_\mathrm{X} &= I_\mathrm{X} n\\ \mathcal{I}_\mathrm{Y} &= I_\mathrm{Y} n\\ \mathcal{I}_\mathrm{Z} &= I_\mathrm{Z} n\end{aligned} \tag{7.83}$$

are the summed rather than the averaged intensities, $\mathbf{I}$ is the 3×3 identity matrix, $\tilde{\mathbf{C}}$ the complex covariance matrix defined in Equation 7.4 on page 53 and $\mathrm{diag}(\cdot)$ denotes the diagonal matrix with diagonal entries specified by $(\cdot)$. Using Equation 7.4 and expanding the right hand side of Equation 7.82 gives

$$\begin{aligned}\Psi(t_1,t_2,t_3) = \Big[&1 - j2\sigma_X^2 t_1 - j2\sigma_Y^2 t_2 - j2\sigma_Z^2 t_3\\ &+ 4(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2)t_1t_2 + 4(c_{XZA}^2 + c_{XZC}^2 - \sigma_X^2\sigma_Z^2)t_1t_3 + 4(c_{YZA}^2 + c_{YZC}^2 - \sigma_Y^2\sigma_Z^2)t_2t_3\\ &+ j8\big(\sigma_X^2\sigma_Y^2\sigma_Z^2 + 2T - (c_{XYA}^2 + c_{XYC}^2)\sigma_Z^2 - (c_{XZA}^2 + c_{XZC}^2)\sigma_Y^2 - (c_{YZA}^2 + c_{YZC}^2)\sigma_X^2\big)t_1t_2t_3\Big]^{-n}\end{aligned} \tag{7.84}$$

where T is defined in Equation 7.8 on page 54.

The joint density is given by the inverse Fourier transformation of the characteristic function [Papoulis 1991]

$$\mathrm{f}(\mathcal{I}_\mathrm{X},\mathcal{I}_\mathrm{Y},\mathcal{I}_\mathrm{Z}) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\big(-j(t_1\mathcal{I}_\mathrm{X} + t_2\mathcal{I}_\mathrm{Y} + t_3\mathcal{I}_\mathrm{Z})\big)\Psi(t_1,t_2,t_3)\,dt_1\,dt_2\,dt_3. \tag{7.85}$$

The characteristic function is of the form

$$\Psi(t_1,t_2,t_3) = (\alpha + \beta t_1)^{-n}, \tag{7.86}$$

where

$$\begin{aligned}\alpha &= a + bt_2\\ \beta &= c + dt_2,\end{aligned} \tag{7.87}$$

and the following parameters depend only on t_3:

$$\begin{aligned}
a &= 1 - j2\sigma_Z^2 t_3 \\
b &= -j2\sigma_Y^2 + 4(c_{YZA}^2 + c_{YZC}^2 - \sigma_Y^2\sigma_Z^2)t_3 \\
c &= -j2\sigma_X^2 + 4(c_{XZA}^2 + c_{XZC}^2 - \sigma_X^2\sigma_Z^2)t_3 \\
d &= 4(c_{XYA}^2 + c_{XYC}^2 - \sigma_X^2\sigma_Y^2) + j8(\sigma_X^2\sigma_Y^2\sigma_Z^2 + 2T - (c_{XYA}^2 + c_{XYC}^2)\sigma_Z^2 \\
&\quad - (c_{XZA}^2 + c_{XZC}^2)\sigma_Y^2 - (c_{YZA}^2 + c_{YZC}^2)\sigma_X^2)t_3.
\end{aligned} \tag{7.88}$$

Using the closed form expression for

$$\int_{-\infty}^{\infty} e^{-jt_1 I_X}\ (\alpha + \beta t_1)^{-n} dt_1 \tag{7.89}$$

(Gradshteyn and Ryzhik [1994], page 365) allows the t_1 integral in Equation 7.85 to be evaluated, giving

$$\mathrm{f}(\mathcal{I}_X, \mathcal{I}_Y, \mathcal{I}_Z) = \frac{j}{8\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(j\mathcal{I}_X)^{n-1}}{(n-1)!(c+dt_2)^n} \exp\left(-j\left(\frac{(a+bt_2)\mathcal{I}_X}{c+dt_2} + t_2\mathcal{I}_Y + t_3\mathcal{I}_Z\right)\right) dt_2 dt_3. \tag{7.90}$$

The integral over t_2 in Equation 7.90 is

$$I = \int_{-\infty}^{\infty} \frac{\exp\left(-j\big((a+bt_2)\mathcal{I}_X/(c+dt_2) + t_2\mathcal{I}_Y\big)\right)}{(c+dt_2)^n} dt_2, \tag{7.91}$$

which can be rewritten in the form

$$I = \exp\left(-j\mathcal{I}_X\frac{b}{d}\right) \int_{-\infty}^{\infty} \frac{\exp(-j\mathcal{I}_Y t_2)\exp\left(-j\big((a-bc/d)\mathcal{I}_X/(c+dt_2)\big)\right)}{(c+dt_2)^n} dt_2. \tag{7.92}$$

Since the argument of the third exponential is bounded, it can be expanded as a power series to give

$$I = \exp\left(-j\mathcal{I}_X\frac{b}{d}\right) \int_{-\infty}^{\infty} \exp(-j\mathcal{I}_Y t_2) \sum_{k=0}^{\infty} \frac{(-j(a-bc/d)\mathcal{I}_X)^k}{k!(c+dt_2)^{n+k}} dt_2. \tag{7.93}$$

Interchanging the order of integration and summation in Equation 7.93 (valid due to uniform convergence of the series) gives

$$I = \exp\left(-j\mathcal{I}_X\frac{b}{d}\right) \sum_{k=0}^{\infty} \frac{(-j(a-bc/d)\mathcal{I}_X)^k}{k!} \int_{-\infty}^{\infty} \frac{\exp(-j\mathcal{I}_Y t_2)}{(c+dt_2)^{n+k}} dt_2, \tag{7.94}$$

and the integral can be evaluated (Gradshteyn and Ryzhik [1994], page 365), to give

$$I = \frac{-j\pi}{d}\left(\frac{-j\mathcal{I}_Y}{d}\right)^{n-1}\exp\left(-j\left(\mathcal{I}_X\frac{b}{d}-\mathcal{I}_Y\frac{c}{d}\right)\right)\sum_{k=0}^{\infty}\frac{\left(-(ad-bc)\mathcal{I}_X\mathcal{I}_Y/d^2\right)^k}{k!(n+k-1)!}. \tag{7.95}$$

The infinite sum can be equated to a Bessel function (Abramowitz and Stegun [1972], page 360) to give

$$I = \frac{-j\pi}{d}\left(\frac{-j\mathcal{I}_Y}{d}\right)^{n-1}\exp\left(-j\left(\mathcal{I}_X\frac{b}{d}-\mathcal{I}_Y\frac{c}{d}\right)\right)\left(\sqrt{\mathcal{I}_X\mathcal{I}_Y(ad-bc)}/d\right)^{2-n}\mathrm{J}_{n-2}\left(2\sqrt{\mathcal{I}_X\mathcal{I}_Y(ad-bc)}/d\right). \tag{7.96}$$

Substituting this result into Equation 7.90 gives

$$\mathrm{f}(\mathcal{I}_X,\mathcal{I}_Y,\mathcal{I}_Z) = \frac{1}{8\pi}\int_{-\infty}^{\infty}\frac{(\mathcal{I}_X\mathcal{I}_Y)^{n/2}\exp\left(-j(b\mathcal{I}_X/d-c\mathcal{I}_Y/d+t_3\mathcal{I}_Z)\right)}{d^2(n-1)!\,(ad-bc)^{(n-2)/2}}\mathrm{J}_{n-2}\left(2\sqrt{\mathcal{I}_X\mathcal{I}_Y(ad-bc)}/d\right)dt_3. \tag{7.97}$$

The integral in Equation 7.97 is difficult to evaluate and the joint density is left in this form. Converting $\mathcal{I}$, the summed intensities (see Equation 7.83 on page 67) into I the averaged intensities, in Equation 7.97 gives

$$\mathrm{f}(I_X,I_Y,I_Z) = \frac{n^3}{8\pi}\int_{-\infty}^{\infty}\frac{(I_XI_Y)^{n/2}\exp\left(-\frac{j}{n}(bI_X/d-cI_Y/d+t_3I_Z)\right)}{n^nd^2(n-1)!\,(ad-bc)^{(n-2)/2}}\mathrm{J}_{n-2}\left(\frac{2\sqrt{I_XI_Y(ad-bc)}}{dn}\right)dt_3. \tag{7.98}$$

7.4 PARAMETER ESTIMATION AND MOMENTS OF THE NEW LIKELIHOOD FUNCTION

The statistics of the radar returns are characterised by the nine elements of the covariance matrix in Equation 7.3. The joint density given in Equation 7.45 (and Equation 7.59) is characterised by the eight parameters a_1, b_1, c_1, d_1, d_2, d_3, d_4 (or $\cos\gamma$) and $|\mathbf{C}|$, which depend through the previous equations on the elements of the covariance matrix. In a practical application of the joint density to SAR, for example in a ML or MAP estimator, these parameters would be unknown and would need to be estimated from intensity data. A method of estimating the eight parameters from intensity data is described here. Note that only seven of the eight parameters are independent. However, it is more convenient to estimate the eight parameters from the data rather than to derive a complicated expression for one in terms of the other seven. The method is based on deriving expressions for these parameters in terms of moments of intensity samples that can be computed from the data. This method was chosen due to speed considerations. Alternatively a maximum likelihood method could be used to estimate the parameters. An optimisation schema would have to be implemented in this case.

Since each intensity value is the average of $2n$ squared samples from a Gaussian PDF, they

are chi-square distributed of order $2n$, as given in Equation 4.7 on page 25, with an expected value equal to twice the variance of the Gaussian [Stark and Woods 2002]. The first moments of the intensities, the expected values, are therefore

$$\begin{aligned} \mu_{\mathrm{I_X}} &= \mathrm{E}[I_{\mathrm{X}}] = 2\sigma_X^2 \\ \mu_{\mathrm{I_Y}} &= \mathrm{E}[I_{\mathrm{Y}}] = 2\sigma_Y^2 \\ \mu_{\mathrm{I_Z}} &= \mathrm{E}[I_{\mathrm{Z}}] = 2\sigma_Z^2 \end{aligned} \tag{7.99}$$

and the variances

$$\begin{aligned} \sigma^2_{\mathrm{I_X}} &= \mathrm{E}[I_{\mathrm{X}} I_{\mathrm{X}}] = \frac{4}{n}\sigma_X^4 \\ \sigma^2_{\mathrm{I_X}} &= \mathrm{E}[I_{\mathrm{Y}} I_{\mathrm{Y}}] = \frac{4}{n}\sigma_Y^4 \\ \sigma^2_{\mathrm{I_X}} &= \mathrm{E}[I_{\mathrm{Z}} I_{\mathrm{Z}}] = \frac{4}{n}\sigma_Z^4. \end{aligned} \tag{7.100}$$

Expressions for the higher-order moments are derived using the characteristic function in Equation 7.84 [Papoulis 1991]. These are the moments of the summed intensities $\mathcal{I}$ as defined in Equation 7.83 rather than the averaged intensities and have to be converted back to the averaged intensities I. The second-order moments are derived as, for example,

$$\mathrm{E}[I_{\mathrm{X}} I_{\mathrm{Y}}] = - \left. \frac{\partial^2 \Psi(t_1, t_2, t_3)}{n^2 \partial t_1 \partial t_2} \right|_{t_1 = t_2 = t_3 = 0} \tag{7.101}$$

where $\Psi(\cdot)$ is given in Equation 7.84. The second-order moments of the intensities, which are contained in the covariance matrix of the intensities $\mathbf{C_I}$, can be derived in terms of the elements of $\mathbf{C}$ using Equation 7.101, giving

$$\mathbf{C_I} = \begin{bmatrix} \sigma^2_{\mathrm{I_X}} & c_{IXY} & c_{IXZ} \\ c_{IXY} & \sigma^2_{\mathrm{I_Y}} & c_{IYZ} \\ c_{IXZ} & c_{IYZ} & \sigma^2_{\mathrm{I_Z}} \end{bmatrix} = \frac{4}{n} \begin{bmatrix} \sigma_X^4 & c^2_{XYA} + c^2_{XYC} & c^2_{XZA} + c^2_{XZC} \\ c^2_{XYA} + c^2_{XYC} & \sigma_Y^4 & c^2_{YZA} + c^2_{YZC} \\ c^2_{XZA} + c^2_{XZC} & c^2_{YZA} + c^2_{YZC} & \sigma_Z^4 \end{bmatrix}. \tag{7.102}$$

The third-order central moment of the intensities m_{IXYZ} can be obtained using

$$\mathrm{E}[I_{\mathrm{X}} I_{\mathrm{Y}} I_{\mathrm{Z}}] = j \left. \frac{\partial^3 \Psi(t_1, t_2, t_3)}{n^3 \partial t_1 \partial t_2 \partial t_3} \right|_{t_1 = t_2 = t_3 = 0} \tag{7.103}$$

and Equation 7.102, giving

$$m_{IXYZ} = \mathrm{E}\left[(I_{\mathrm{X}} - \mathrm{E}[I_{\mathrm{X}}])(I_{\mathrm{Y}} - \mathrm{E}[I_{\mathrm{Y}}])(I_{\mathrm{Z}} - \mathrm{E}[I_{\mathrm{Z}}])\right] = \frac{16T}{n^2}. \tag{7.104}$$

Using Equations 7.7 and 7.102, the determinant, $|\mathbf{C}|$, can be written in terms of the moments as

$$|\mathbf{C}|^{1/2} = \frac{1}{8}\Big(n\mu_{\mathrm{I_X}} c_{IYZ} + n\mu_{\mathrm{I_Y}} c_{IXZ} + n\mu_{\mathrm{I_Z}} c_{IXY} - \mu_{\mathrm{I_X}}\mu_{\mathrm{I_Y}}\mu_{\mathrm{I_Z}} - n^2 m_{IXYZ}\Big). \tag{7.105}$$

Relatively straightforward manipulations of Equations 7.6-7.8 and Equations 7.102-7.105 allow the remaining seven parameters to be written in terms of the moments as

$$
\begin{aligned}
a_1 &= \frac{nc_{\mathbf{I}YZ} - \mu_{\mathrm{I_Y}}\mu_{\mathrm{I_Z}}}{4\,|\mathbf{C}|^{1/2}} \\
b_1 &= \frac{nc_{\mathbf{I}XZ} - \mu_{\mathrm{I_X}}\mu_{\mathrm{I_Z}}}{4\,|\mathbf{C}|^{1/2}} \\
c_1 &= \frac{nc_{\mathbf{I}XY} - \mu_{\mathrm{I_X}}\mu_{\mathrm{I_Y}}}{4\,|\mathbf{C}|^{1/2}} \\
d_1^2 &= \frac{n\mu_{\mathrm{I_Z}}^2 c_{\mathbf{I}XY} + n^2 c_{\mathbf{I}XZ}c_{\mathbf{I}YZ} - n^2\mu_{\mathrm{I_Z}}m_{\mathbf{I}XYZ}}{16\,|\mathbf{C}|} \\
d_2^2 &= \frac{n\mu_{\mathrm{I_X}}^2 c_{\mathbf{I}YZ} + n^2 c_{\mathbf{I}XY}c_{\mathbf{I}XZ} - n^2\mu_{\mathrm{I_X}}m_{\mathbf{I}XYZ}}{16\,|\mathbf{C}|} \\
d_3^2 &= \frac{n\mu_{\mathrm{I_Y}}^2 c_{\mathbf{I}XZ} + n^2 c_{\mathbf{I}XY}c_{\mathbf{I}YZ} - n^2\mu_{\mathrm{I_Y}}m_{\mathbf{I}XYZ}}{16\,|\mathbf{C}|} \\
d_4 &= \frac{1}{128\,|\mathbf{C}|^{3/2}}\Big[- n^4 m_{\mathbf{I}XYZ}^2 + n^2 m_{\mathbf{I}XYZ}\left(\mu_{\mathrm{I_X}}\mu_{\mathrm{I_Y}}\mu_{\mathrm{I_Z}} + n\mu_{\mathrm{I_X}}c_{\mathbf{I}YZ} + n\mu_{\mathrm{I_Y}}c_{\mathbf{I}XZ} + n\mu_{\mathrm{I_Z}}c_{\mathbf{I}XY}\right) \\
&\quad +2n^2\left(nc_{\mathbf{I}XY}c_{\mathbf{I}XZ}c_{\mathbf{I}YZ} - \mu_{\mathrm{I_X}}\mu_{\mathrm{I_Y}}c_{\mathbf{I}XZ}c_{\mathbf{I}YZ} - \mu_{\mathrm{I_X}}\mu_{\mathrm{I_Z}}c_{\mathbf{I}XY}c_{\mathbf{I}YZ} - \mu_{\mathrm{I_Y}}\mu_{\mathrm{I_Z}}c_{\mathbf{I}XY}c_{\mathbf{I}XZ}\right)\Big].
\end{aligned}
\tag{7.106}
$$

Equations 7.105 and 7.106 therefore allow the parameters of the joint density $\mathrm{f}(I_\mathrm{X}, I_\mathrm{Y}, I_\mathrm{Z})$ to be estimated from intensity data by substituting sample moments for the theoretical moments.

7.5 CONCLUSION

In this chapter a closed-form expression for the joint density of the intensity data from a three-polarisation n-look SAR system has been derived. This joint density describes the averaged intensity values from n samples from a complex three-dimensional zero-mean circular Gaussian random process. Two different cases have to be distinguished in the derivation. In the first case with $n > 1$, results provided by Miller et al. [1958] are used to solve the integrals. In the second case for $n = 1$ the derivation of Miller et al. [1958] breaks down and an alternative way of solving the integral is given. The resulting expressions are in the form of an infinite series of modified Bessel functions with seven independent parameters. The convergence of this expression is rapid as will be shown in Chapter 9.

In order to theoretically validate the result presented here, three special cases of this density have been investigated. For these cases the joint density simplifies to known expressions:

- **The uncorrelated case:** The three intensity values are uncorrelated from each other (the covariance matrix is a diagonal matrix). The joint density collapses to three independent chi-square distributions of the order $2n$ as expected.

- **The two-dimensional case:** One intensity value is independent of the other two. The remaining two are correlated to each other. The joint density simplifies to the product of a chi-

square density and a two-dimensional PDF equivalent to the density of the two-polarisation n-look SAR data derived by Lee et al. [1994b].

- **The real case:** There are no correlations between the imaginary and real parts of the complex signals. Removing these correlations makes this case equivalent to the sum of $2n$ intensity samples from a three-dimensional *real* zero-mean Gaussian distribution. Miller et al. [1958] derived the PDF for this case which matches the one given here.

The characteristic function of the density has been derived. The characteristic function provides a second approach to forming an expression for the joint density under investigation. This approach was followed through until only one integral was left which was considered too difficult to solve. However this expression for the joint density offers an alternative way of numerically evaluating the joint density.

The parameters of the joint density are defined as functions of the covariance matrix of the circular Gaussian distribution. A method was developed to estimate these parameters from intensity sample data. The parameters are expressed in terms of the moments of the intensity data with the help of the characteristic function which provides an easy way of deriving expressions for the moments of the joint density.

The new likelihood function is analysed and numerically validated in Chapter 9. Simulated and true data is matched against the new likelihood function to confirm its validity. The method of estimating parameters from sample data is assessed.

Chapter 8

PARAMETER ESTIMATION FOR THE GRF MODEL

In this Chapter a new method [Hagedorn et al. 2001] of estimating the GRF model parameters from a given sample data set is presented. The sample data set is called the 'data set'.

The method presented is based upon the minimisation of a cost function with regard to the model parameters. The cost function gives a value for the difference between the local model PDF and a PDF estimated from the data set. Both PDFs are discrete and have a set number of states. It is shown that the model does not distinguish between some local configurations and these can be combined into a single state which is called here an 'equivalency'. The equivalencies reduce the complexity of the model by a large amount and make the comparison between model and data practical. Note that equivalencies have been used before by Chen [1988]. He gives a brief definition of them and an example for the eight-neighbourhood. In this thesis the total number of equivalencies is derived for any given neighbourhood and clique configuration. A method of remapping the discrete local GRF PDF with regard to the equivalencies is also presented.

In Section 8.1 the local conditional PDF is discussed in detail. In Section 8.2 equivalencies are defined. A method of ordering and remapping the equivalencies and some examples are given. In Section 8.3 the cost function is derived and in Section 8.4 conclusions are drawn. Practical experiments and results are given in Chapter 10.

8.1 THE LOCAL CONDITIONAL PDF

In this section the local conditional PDF of the GRF model is analysed. A different notation in vector form of the potential function leads into the definition of equivalencies presented in the next section. For more detail of the MRF stochastic model in general and the notation used here the reader is referred to Sections 5.7 to 5.9.

Following the development given in Section 5.9, the local conditional PDF for the GRF is given by

$$\Pi(x_i|\mathbf{x}_{\mathcal{N}_i}) = \frac{1}{Z}\exp\left[-\sum_{\forall c\in\mathcal{C}_i} V_c(x_i,\mathbf{x}_{\mathcal{N}_i})\right], \tag{8.1}$$

where

$$Z = \sum_{\forall x_i' \in \mathbf{X}_i} \exp\left[- \sum_{\forall c \in \mathcal{C}_i} V_c(x_i', \mathbf{x}_{\mathcal{N}_i}) \right]. \tag{8.2}$$

The potential function is given by

$$V_c(x_i, \mathbf{x}_{\mathcal{N}_i}) = \begin{cases} -\beta_C & \text{if all } x \in \mathbf{x}_c \text{ are equal} \\ +\beta_C & \text{otherwise} \end{cases} \quad c \in C \in \mathcal{C}, |c| \geq 2, \tag{8.3}$$

where β_C is a constant with C indicating the clique type and $\mathbf{x}_c$ are the site values of clique c where c is a clique out of all possible cliques in the local neighbourhood $(i, \mathcal{N}_i)$. Each clique type therefore has a β value assigned to it. Further note that the potential functions assign a value to what will be called a local configuration $(x_i, \mathbf{x}_{\mathcal{N}_i})$. The total number of possible local configurations depends on the number of sites that are in the configuration and on the number of classes (or valid label values). This is discussed in detail further on.

First, new notation is introduced which simplifies the interpretation of the local conditional PDF of the GRF in Equation 8.1. The sum in the exponential of the potential function in Equation 8.1 can be replaced by the vector product

$$\sum_{\forall c \in \mathcal{C}_i} V_c(x_i, \mathbf{x}_{\mathcal{N}_i}) = \mathbf{w}(x_i, \mathbf{x}_{\mathcal{N}_i})\Theta \tag{8.4}$$

where Θ is a column vector of parameter values defined as, for example, $\Theta = [\beta_1 \ \beta_2]^T$ for the four-neighbourhood and $\Theta = [\beta_1 \ \beta_2 \ \beta_3 \ \beta_4]^T$ for the eight-neighbourhood. Each β corresponds to a clique type. The βs are ordered as follows: $\Theta = [\ |\ -]^T$ for the four-neighbourhood and $\Theta = [\ |\ -\ \backslash\ /\ ]^T$ for the eight-neighbourhood. Thus β_1 corresponds to the vertical clique type and correspondingly for the rest.

The row vector $\mathbf{w}(x_i, \mathbf{x}_{\mathcal{N}_i})$ is defined as $\mathbf{w}(x_i, \mathbf{x}_{\mathcal{N}_i}) = [w_1(x_i, \mathbf{x}_{\mathcal{N}_i}) \ \ w_2(x_i, \mathbf{x}_{\mathcal{N}_i}) \ \ \ldots]$ in which the integer value

$$w_k(x_i, \mathbf{x}_{\mathcal{N}_i}) = \sum_{c \in C_k} w_c \tag{8.5}$$

where

$$w_c = \begin{cases} -1 & \text{, if all } x \in \mathbf{x}_c \text{ are equal,} \\ +1 & \text{, otherwise,} \end{cases} \tag{8.6}$$

and where C_k is the kth clique type.

For example Figure 8.1 shows the two cliques in the four-neighbourhood of the horizontal clique type. Figure 8.2 shows three example configurations of $(x_i, \mathbf{x}_{\mathcal{N}_i})$ for the four-neighbourhood and three classes. The different gray shades indicate different site values. For the examples in Figure 8.2 w_2, which corresponds to the horizontal clique type, is from left to right $w_2 = -2$, $w_2 = 2$ and $w_2 = 0$ respectively (the complete vectors are $\mathbf{w} = [2 \ \ -2]$, $\mathbf{w} = [-2 \ \ \ 2]$ and $\mathbf{w} = [2 \ \ \ 0]$ respectively).

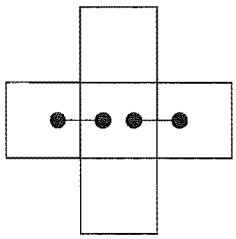


Figure 8.1 The two horizontal cliques in the four-neighbourhood.

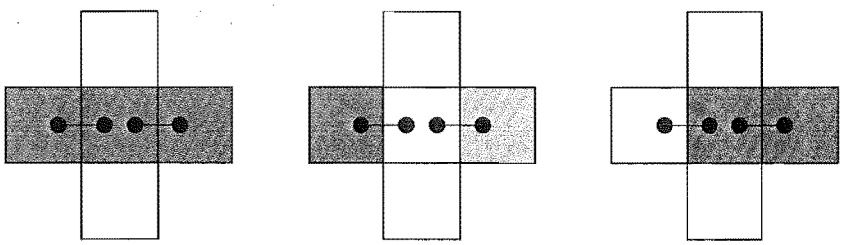


Figure 8.2 Example configurations in the four-neighbourhood for the two horizontal cliques.

Equation 8.1 can now be rewritten as

$$\Pi(x_i|\mathbf{x}_{\mathcal{N}_i}) = \frac{\exp\left[-\mathbf{w}(x_i, \mathbf{x}_{\mathcal{N}_i})\Theta\right]}{\sum_{\forall x_i' \in \mathbf{X}_i} \exp\left[-\mathbf{w}(x_i', \mathbf{x}_{\mathcal{N}_i})\Theta\right]}. \tag{8.7}$$

8.2 EQUIVALENCIES

Let L be the number of classes present in a label image ($\mathbf{X}_i = \{1, 2, \ldots, L\}$) and N the number of sites in a neighbourhood system (for example $N = 4$ for the four- and $N = 8$ for the eight-neighbourhood). A local configuration $(x_i, \mathbf{x}_{\mathcal{N}_i})$ therefore consists of $N+1$ sites and L^{N+1} different configurations exist. It is clear that many of these configurations are represented by the same $\mathbf{w}$ vector. The group of configurations represented by a single unique $\mathbf{w}$ vector is referred to as an equivalency. The individual configurations in an equivalency are indistinguishable from the model's perspective. One configuration produces the same conditional probability (see Equation 8.7) as any other configuration in the same equivalency independent of the parameter vector Θ.

Let ξ_k be the number of cliques of type k in a local configuration $(x_i, \mathbf{x}_{\mathcal{N}_i})$, and ζ be the number of different clique types ($k \in \{1, \ldots, \zeta\}$). Due to the definition of w_k in Equation 8.6 there are always $\xi_k + 1$ possible different values for w_k and the vector $\mathbf{w}$ has ζ elements. For $\xi_k = 2$ (for example there are two cliques of the horizontal clique type in the four-neighbourhood as shown in Figure 8.2), w_k may take the values -2, 0, 2, for example. The number of equivalencies can therefore be calculated to be given by

$$\text{Number of equivalencies} = \prod_{k=1}^{\zeta} (\xi_k + 1). \tag{8.8}$$

If only pairwise cliques are considered, then $\xi_k = 2$ and Equation 8.8 can be simplified to 3^ζ.

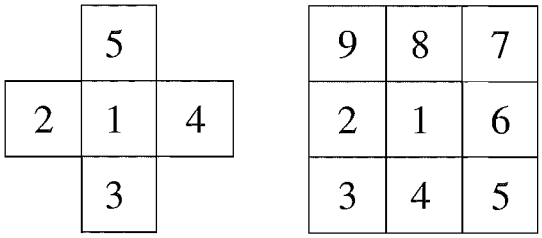


Figure 8.3 The order of the sites in the local configuration $\hat{\mathbf{x}}_j$ for the four- and eight-neighbourhood.

This reduces the dimensionality of the configurations space by a large amount; in the case of pairwise cliques only 3^ζ equivalencies have to be considered instead of L^{N+1} different configurations. Note the number of equivalencies is independent of the number of classes L.

Fixing the β values for different clique types to one single value can be interpreted as merging the corresponding sets of clique types. This results in ξ being the combined number of the two clique types and reducing the total number of clique types ζ. This results in a further reduction of the number of equivalencies. Such a simplification may be justified when, for example, the prior knowledge exists that there are no systematic differences between horizontal and vertical features. In this case the horizontal clique and the vertical clique each have the same β value. Additionally in the eight-neighbourhood case the two diagonal cliques may be assigned to one β value.

8.2.1 Remapping the equivalencies

In order to simplify the PDF and the comparisons of the PDF, a mapping is performed to take account of the equivalencies, as follows. Let $\hat{\mathbf{x}}$ denote the local configuration $(x_i, \mathbf{x}_{\mathcal{N}_i})$ and $\hat{\mathbf{x}}_j$ the jth realisation of this configuration. Define an ordering for the configurations based on the $(N+1)$-tuple of the site values. For example the four-neighbourhood $\hat{\mathbf{x}}_0 = (0,0,0,0,0)$, where the first element in the tuple is the value of the site in the center of the configuration, x_i. The order of the remaining elements in the tuple is indicated for the four- and eight-neighbourhood in Figure 8.3. To complete the ordering, the site values in the tuple are interpreted as a number with base L. For data with binary labelling therefore each possible local configuration is represented by a unique binary number and j is the corresponding decimal number. Thus for the four-neighbourhood, configuration $\hat{\mathbf{x}}_6 = (0,0,1,1,0)$, etc. Table 8.1 gives two examples for the four-neighbourhood, one for the binary case $L = 2$ and one for $L = 3$. The ordering and numbering of these configurations can be done arbitrarily as long as it is done consistently.

In the paragraphs below the local conditional PDF and the corresponding PDF of the equivalencies are expressed in form of vectors called $\mathbf{\Pi}$ and $\mathbf{\Psi}$ respectively. Then a matrix $\mathbf{A}$ is defined such that $\mathbf{\Psi} = \mathbf{A}\,\mathbf{\Pi}$. This gives a practical approach to the remapping of the equivalencies and is used in the next section in the development of the cost function.

Since the local conditional PDF (c.f. Equation 8.1), $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$ is a discrete function over the

	$L=2$	$L=3$
$\hat{\mathbf{x}}_0$	$(0,0,0,0,0)$	$(0,0,0,0,0)$
$\hat{\mathbf{x}}_1$	$(0,0,0,0,1)$	$(0,0,0,0,1)$
$\hat{\mathbf{x}}_2$	$(0,0,0,1,0)$	$(0,0,0,0,2)$
$\hat{\mathbf{x}}_3$	$(0,0,0,1,1)$	$(0,0,0,1,0)$
$\hat{\mathbf{x}}_4$	$(0,0,1,0,0)$	$(0,0,0,1,1)$
$\hat{\mathbf{x}}_5$	$(0,0,1,0,1)$	$(0,0,0,1,2)$
$\hat{\mathbf{x}}_6$	$(0,0,1,1,0)$	$(0,0,0,2,0)$
$\vdots$	$\vdots$	$\vdots$

Table 8.1 Example for the numbering schema of the configurations $\hat{\mathbf{x}}_j$ (four-neighbourhood). The sites in the configurations have label values chosen from the set $\{1,\ldots,L\}$. The j number is generated by interpreting the vector of label values as a number with the base system L. The order of the site values in the vector is given in Figure 8.3.

local configuration space $(x_i, \mathbf{x}_{\mathcal{N}_i})$, it can be seen as a vector

$$\mathbf{\Pi} = [\Pi(\hat{\mathbf{x}}_0) \;\; \Pi(\hat{\mathbf{x}}_1) \;\; \ldots]^T \tag{8.9}$$

with a length L^{N+1}. Let $\mathbf{\Psi} = [\Psi_1 \;\; \Psi_2 \;\; \ldots]^T$ be the corresponding vector of equivalencies of length $\prod_{k=1}^{\zeta}(\xi_k + 1)$ (as defined in Equation 8.8) and $\{\hat{\mathbf{x}}_{m_1}, \hat{\mathbf{x}}_{m_2}, \ldots, \hat{\mathbf{x}}_{m_\mu}\}$ be the set of μ local configurations that make the mth equivalency. Due to the definition of the equivalency,

$$\mathbf{w}(\hat{\mathbf{x}}_{m_1}) = \mathbf{w}(\hat{\mathbf{x}}_{m_2}) = \cdots = \mathbf{w}(\hat{\mathbf{x}}_{m_\mu}) \tag{8.10}$$

is valid and the corresponding value of the equivalency is defined as

$$\Psi_m = \Pi(\hat{\mathbf{x}}_{m_1}) + \Pi(\hat{\mathbf{x}}_{m_2}) + \cdots + \Pi(\hat{\mathbf{x}}_{m_\mu}). \tag{8.11}$$

The ordering of the equivalencies can be done arbitrarily. The forming of the equivalencies out of the discrete PDF is referred to as a remapping operation and can be given in the form

$$\mathbf{\Psi} = \mathbf{A}\,\mathbf{\Pi} \tag{8.12}$$

where $\mathbf{A}$ is a remapping matrix of the size $\left[L^{N+1}\right] \times \left[\prod_{k=1}^{\zeta}(\xi_k+1)\right]$. The remapping matrix consists of only zeros and ones. Each column has a single one in it, the rest of the elements in that column are zero. Each row corresponds to one equivalency and has ones at the positions $m_1, m_2, \ldots m_\mu$, the elements of the $\mathbf{\Pi}$ vector that contribute to this equivalency.

8.2.2 Examples

Table 8.2 shows examples of different parameter space setups and the corresponding ξ, ζ and equivalency numbers. Depending on the number of classes L there are a large number of local configurations. For example with $L = 10$ and the eight-neighbourhood there are 10^9 different configurations and only 81 equivalencies if every clique type is assigned a different β value. This

Four-neighbourhood			
Θ	ξ	ζ	Number of equivalencies
$[\mid\ -]^T$	2	2	9
$\{\mid\ -\}$	4	1	5
Eight-neighbourhood			
Θ	ξ	ζ	Number of equivalencies
$[\mid\ -\ \diagdown\ \diagup]^T$	2	4	81
$[\{\mid\ -\}\ \{\diagdown\ \diagup\}]^T$	4	2	25
$\{\mid\ -\ \diagdown\ \diagup\}$	8	1	9

Table 8.2 Examples of parameter space setups for the four- and eight-neighbourhood and the corresponding ξ, ζ and equivalency numbers. $\mid$ indicates the vertical clique type and correspondingly $-$, $\diagdown$ and $\diagup$. Further, for example, $\{\mid\ -\}$ indicates that the vertical and the horizontal clique types are assigned the same β value.

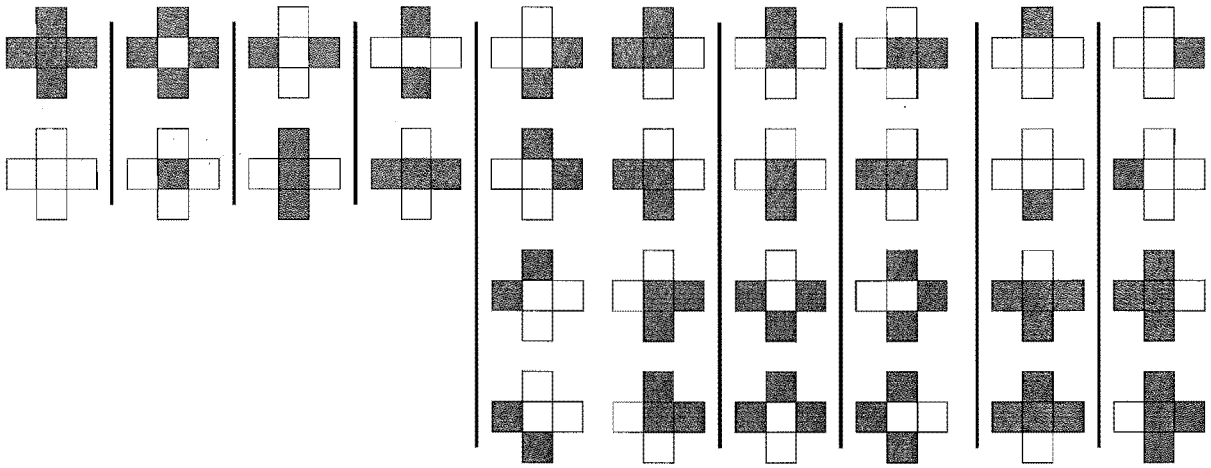


Figure 8.4 All possible local configurations for a two class four-neighbourhood system, arranged in groups with identical **w**. In this case $\Theta = [\beta_1\ \ \beta_2]$ with $\beta_1 \neq \beta_2$. The **w** vector from left to right: $[-2\ -2]$, $[2\ \ 2]$, $[2\ -2]$, $[-2\ \ 2]$, $[0\ \ 0]$, $[0\ \ 2]$, $[2\ \ 0]$, $[0\ -2]$ and $[-2\ \ 0]$.

illustrates how the selection of the potential function can reduce the system complexity by a large amount.

Figures 8.4 and 8.5 show all possible configurations for the four-neighbourhood with two classes $L = 2$ (binary label image). The configurations are grouped in equivalencies. In Figure 8.4 both the horizontal and the vertical clique are assigned independent β values while in Figure 8.5 both clique types are fixed to the same β value. Note that in the binary case configurations and their inverted configurations are always in the same group.

8.3 THE COST FUNCTION

In this section a cost function (error function) is developed which describes the difference between the GRF model and a specific data set. This cost function is then minimised with regard to the model parameters Θ and the set of parameters which produces the smallest value in the cost

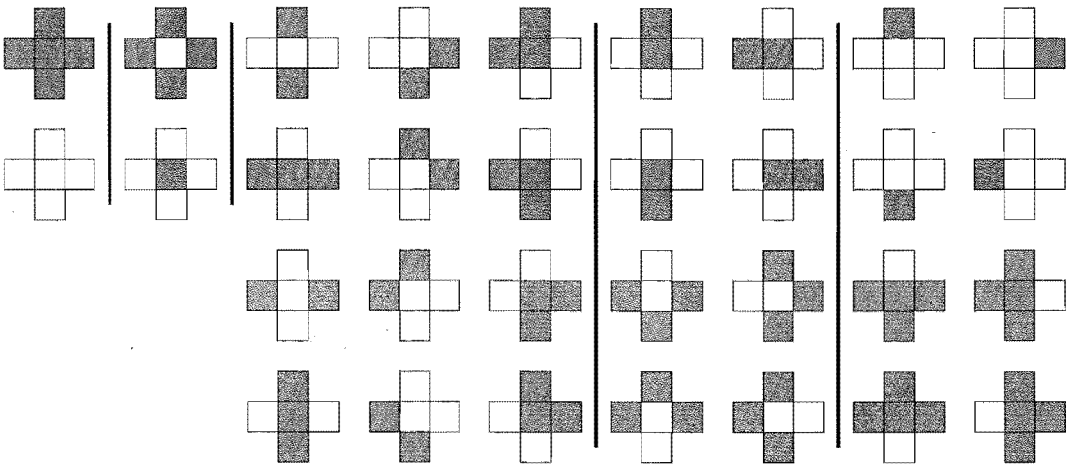


Figure 8.5 All possible local configurations for a two class four-neighbourhood system, arranged in groups with identical $\mathbf{w}$. In this case $\Theta = [\beta_1 \ \beta_2] = \beta$ with $\beta_1 = \beta_2$. $\mathbf{w}$ from left to right: -4, 4, 0, 2 and -2.

function is considered optimal.

Ideally the local joint model PDF of the GRF $\Pi(x_i, \mathbf{x}_{\mathcal{N}_i})$ would be compared to an estimated local joint PDF $\hat{\Pi}(x_i, \mathbf{x}_{\mathcal{N}_i})$ from given data. $\hat{\Pi}(x_i, \mathbf{x}_{\mathcal{N}_i})$ can be calculated from the relative frequency of independent local configuration samples from the data. It is however impractical to compute $\Pi(x_i, \mathbf{x}_{\mathcal{N}_i})$ since this can only be done by evaluating the partitioning function of the GRF model. The local conditional PDF $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$ however is fast and convenient to compute. The same density can be estimated from the data set and parameters can be estimated by minimising the difference between the two densities. This cost function can be expressed in the following form. Let $\mathbf{\Psi}^{(x_i|\mathbf{x}_{\mathcal{N}_i})} = \mathbf{A}\,\mathbf{\Pi}^{(x_i|\mathbf{x}_{\mathcal{N}_i})}$ be the discrete vector of the local conditional GRF after the equivalencies are remapped with the appropriate $\mathbf{A}$ and let $\hat{\mathbf{\Psi}}^{(x_i|\mathbf{x}_{\mathcal{N}_i})}$ be the same vector where the values have been estimated from the data. The cost function for the local conditional densities $\mathcal{E}_{\text{cond}}(\Theta)$ can then be written as

$$\mathcal{E}_{\text{cond}}(\Theta) = \sum_l \left[\Psi_l^{(x_i|\mathbf{x}_{\mathcal{N}_i})}(\Theta) - \hat{\Psi}_l^{(x_i|\mathbf{x}_{\mathcal{N}_i})}\right]^2. \tag{8.13}$$

Note however that in this case two local conditional densities are compared and due to the fact that they are conditional PDFs, large probabilities can arise for x_i given $\mathbf{x}_{\mathcal{N}_i}$ even if the neighbourhood configuration $\mathbf{x}_{\mathcal{N}_i}$ hardly ever occurs. This is likely to place high importance on differences that are not important at all and would guide the optimisation in the wrong direction.

To avoid the problem identified in the last paragraph the following cost function was developed

$$\mathcal{E}(\Theta) = \sum_l \left[\tilde{\Psi}_l^{(x_i,\mathbf{x}_{\mathcal{N}_i})}(\Theta) - \hat{\Psi}_l^{(x_i,\mathbf{x}_{\mathcal{N}_i})}\right]^2. \tag{8.14}$$

where $\tilde{\Psi}_l^{(x_i,\mathbf{x}_{\mathcal{N}_i})}(\Theta)$ is defined by

$$\tilde{\mathbf{\Psi}}^{(x_i,\mathbf{x}_{\mathcal{N}_i})}(\Theta) = \mathbf{A}\left(\mathbf{\Pi}^{(x_i|\mathbf{x}_{\mathcal{N}_i})}(\Theta)\left(\hat{\mathbf{\Pi}}^{(\mathbf{x}_{\mathcal{N}_i})} + \varepsilon\right)\right). \tag{8.15}$$

An estimate of the local joint density $\Pi(x_i, \mathbf{x}_{\mathcal{N}_i})$ is obtained by multiplying the local conditional density $\Pi(x_i|\mathbf{x}_{\mathcal{N}_i})$ with an estimate from the data of the joint density of the neighbourhood $\hat{\Pi}(\mathbf{x}_{\mathcal{N}_i})$. Since $\hat{\Pi}(\mathbf{x}_{\mathcal{N}_i})$ is an estimate from a limited number of samples, a small value ε is added to avoid it ever becoming zero. Then the equivalencies are remapped. Note that good results were obtained using $\varepsilon = 1/(n_{\text{equi}}L^{N+1})$ where n_{equi} is the number of equivalencies.

The set of parameters Θ can now be estimated by minimising the cost function in Equation 8.14 as follows

$$\hat{\Theta} = \arg\min_{\Theta} \mathcal{E}(\Theta). \tag{8.16}$$

Experimental results of this new method of estimating the parameters of a GRF model are presented in Chapter 10. An alternative metric to the one presented in Equation 8.14 based on the same discrete PDF's could be used.

8.4 CONCLUSIONS

In this Chapter a novel way of estimating parameters for the GRF model from a given sample data set is derived. For this the idea of equivalencies is introduced. The equivalencies reduce the GRF model to its true complexity and gives an insight into the behaviour of the model. It is shown that the model does not distinguish between local configurations out of the same family of configurations belonging to an equivalency. These configurations can be treated as one and the total number of equivalencies is independent of the number of classes used in the model. The discrete GRF model is now expressed in vector form which can be reduced greatly by help of the equivalencies.

The cost function derived is based upon the local joint PDF of the model and its equivalent estimated from the data. The local joint PDF of the model however is impractical to estimate, due to the complexity of the partitioning function, and is therefore derived from the local conditional PDF of the model and the local joint PDF of the neighbourhood system as estimated from the data set. The resulting local joint PDF of the model and the data are both remapped and greatly reduced with regard to the equivalencies and the squared difference between both is calculated. This cost value can be minimised with regard to the model parameters which will ensure that the model fits the data as well as possible. Results of this method are given in Chapter 10.

Part III

Contribution: Simulation and application

Chapter 9

VALIDATION OF THE NEW LIKELIHOOD FUNCTION FOR POLARIMETRIC SAR DATA AND THE PARAMETER ESTIMATION

In this chapter the new likelihood function, which has been derived in Chapter 7, is analysed and validated. The main results of Chapter 7 are the two different forms of the joint density for multilook polarimetric intensity SAR data. Equation 7.45 on page 60 gives this joint density for the number of looks greater than one ($n > 1$) and Equation 7.59 on page 62 for the number of looks equal to one ($n = 1$). Each of the individual intensity signals is known to be chi-square distributed (see Section 4.1.3 on page 25). In a monostatic polarimetric SAR system the three intensity signals are correlated and therefore the joint density can be seen as a correlated three dimensional chi-square distribution. The correlation arises from the underlying three dimensional complex signal, which has a zero-mean circular Gaussian density. The complex covariance matrix of such a system has been shown to have a complex Wishart density (see Section 4.1.4.1 on page 28). Different types of correlated multidimensional chi-square distributions have been derived, for examples see Kotz et al. [1972]. The PDF derived in Chapter 7 however has not been presented before.

The remainder of this Chapter is organised as follows: First in Section 9.1 a proof for convergence of the infinite sum term in the joint density is given. In Section 9.2 the joint density is compared to simulated data and it is shown that it describes this data accurately. Then in Section 9.3 the method introduced in Section 7.4 for estimating the parameters in the joint density is discussed and examples are given. In Section 9.4 real SAR data is compared to the new PDF and finally in Section 9.5 discussions and conclusions concerning this chapter are presented.

9.1 CONVERGENCE OF THE INFINITE SUM TERM IN THE NEW LIKELIHOOD FUNCTION

Equation 7.45 on page 60 gives the joint density $\mathrm{f}(I_{\mathrm{X}}, I_{\mathrm{Y}}, I_{\mathrm{Z}})$ in the form of a constant multiplied by a negative exponential containing the arguments $I_{\mathrm{X}}, I_{\mathrm{Y}}$ and I_{Z}, and this is in turn multiplied by an infinite sum of terms involving three modified Bessel functions each containing combinations of the function arguments. This section shows that the infinite sum term converges.

A simplified form of the sum term is

$$\sum_{k=n-1}^{\infty} k(-1)^{k-n+1} C_{k-n+1}^{n-1}(\cos\gamma) \mathrm{I}_k(K_1)\,\mathrm{I}_k(K_2)\,\mathrm{I}_k(K_3) = \sum_{k=n-1}^{\infty} B_k \tag{9.1}$$

where K_1, K_2 and K_3 are real positive numbers. The modulus of B_k is

$$|B_k| = k\left|C_{k-n+1}^{n-1}(\cos\gamma)\right| \left|\mathrm{I}_k(K_1)\right| \left|\mathrm{I}_k(K_2)\right| \left|\mathrm{I}_k(K_3)\right|. \tag{9.2}$$

An upper bound for the modified Bessel functions is given by (see Abramowitz and Stegun [1972] page 362)

$$|\mathrm{I}_k(K)| \leq \frac{\left|\frac{1}{2}K\right|^k \exp\left(\Re(K)\right)}{k!} \tag{9.3}$$

($\Re(K)$ is the real part of K, which is K in this case) and for the Gegenbauer function (see Abramowitz and Stegun [1972] page 786)

$$\left|C_{k-n+1}^{n-1}(\cos\gamma)\right| \leq \binom{k+n-2}{k-n+1} = \frac{(k+n-2)!}{(k-n+1)!(2n-3)!}. \tag{9.4}$$

With Equations 9.3 and 9.4 the upper bound for $|B_k|$ can be obtained as

$$|B_k| \leq \frac{k\,(k+n-2)!}{(k-n+1)!(2n-3)!} \frac{\exp(K_1+K_2+K_3)\left(\frac{1}{8}K_1K_2K_3\right)^k}{(k!)^3}. \tag{9.5}$$

Substituting $K_4 = \exp(K_1+K_2+K_3)/(2n-3)!$ and $K_5 = \frac{1}{8}K_1K_2K_3$ and expanding Equation 9.5 results in

$$|B_k| \leq \frac{(k+n-2)!K_4K_5^k}{(k-n+1)!(k!)^2(k-1)!} = U_k. \tag{9.6}$$

Now consider the upper bound for large values of k. Using Stirling's formula [Abramowitz and Stegun 1972],

$$k! \sim e^{-(k+1)}(k+1)^{k+1/2}\sqrt{2\pi}, \tag{9.7}$$

gives

$$U_k \sim \frac{K_4\,K_5^k\,(k+n-1)^{k+n-3/2}\,e^{k-n+2}\,e^{2(k+1)}\,e^k}{\left(\sqrt{2\pi}\right)^3\,(k-n+2)^{k-n+3/2}\,(k+1)^{2k+1}\,k^{k-1/2}\,e^{k+n-1}} = V_k. \tag{9.8}$$

Again, for large values of k the following inequalities hold true

$$\begin{aligned} (k+n-1)^{k+n-3/2} &\leq (2k)^{2k} \\ (k-n+2)^{k-n+3/2} &\geq 1 \\ (k+1)^{2k+1}k^{k-1/2} &\geq k^{3k} \end{aligned} \tag{9.9}$$

and V_k can be written

$$\begin{aligned} V_k &= \frac{K_4 K_5^k \, e^{3k-2n+5}}{\left(\sqrt{2\pi}\right)^3} \times \frac{(k+n-1)^{k+n-3/2}}{(k-n+2)^{k-n+3/2}\,(k+1)^{2k+1}\,k^{k-1/2}} \\ &\leq \frac{K_4 K_5^k \, e^{3k-2n+5}}{\left(\sqrt{2\pi}\right)^3} \times \frac{(2k)^{2k}}{k^{3k}} \\ &= \frac{K_4 e^{-2n+5}}{\left(\sqrt{2\pi}\right)^3} \times \left(\frac{4K_5 e^3}{k}\right)^k = K_6 \left(\frac{K_7}{k}\right)^k \end{aligned} \tag{9.10}$$

where K_6 and K_7 are some positive constants. $(K_7/k)^k$ decays to zero faster than r^k for any $0 < r < 1$. This proves the convergence of the sum in Equation 7.45 on page 60 by the comparison test.

9.2 COMPARING THE NEW LIKELIHOOD FUNCTION WITH SIMULATED DATA

In this Section it is shown that simulated sample data fits the new PDF derived in Chapter 7. The expressions for the new likelihood function in Equation 7.45 on page 60 and Equation 7.59 on page 62 are validated by generating a large number of random numbers drawn from the exact distribution, calculating histograms and comparing them with the density function calculated using Equations 7.45 and 7.59. Results are presented in Figures 9.1 to 9.6 and Table 9.2.

Three different covariance matrices were used to generate the simulated data. Covariance matrix elements for $\mathbf{C}_1$ were estimated from an actual polarimetric SAR image and additionally two artificial covariance matrices $\mathbf{C}_2$ and $\mathbf{C}_3$ were also used. The values for all three covariance matrices are given in Table 9.1.

Using the given parameters, a data set comprising of three pairs of real and imaginary parts as samples from a six-dimensional Gaussian distribution was created (see Equation 7.10 on page 54 for the exact PDF). Intensity samples I_X, I_Y and I_Z were formed from these for a given number of looks n as shown in Equation 7.12 on page 55. In the first case, with $\mathbf{C}_1$, the number of looks is $n = 1$ in the second case with $\mathbf{C}_2$ $n = 3$ and in the third case with $\mathbf{C}_3$ $n = 4$. In all three cases 10^8 intensity samples were generated and then binned into three-dimensional histograms. Three one-dimensional cuts through the histograms are shown in Figures 9.1 - 9.3. Two-dimensional slices of the histograms are shown in Figures 9.4(b), 9.5(b) and 9.6(b). For comparison, the PDF was calculated from the expressions in Equation 7.45 ($n > 1$) and Equation 7.59 ($n = 1$) for the same covariance values. The corresponding parameter values for the covariance matrices were calculated using Equations 7.6 - 7.9 and are listed as 'actual' parameter values in Table 9.3. These parameters were then used to compute the PDF which was compared to the histograms. In Figures 9.1 - 9.3 the PDF is plotted against I_X for the three different sets of parameters. The curves can be seen to closely match the histogram. Likewise the two-dimensional plots of the PDF shown in Figures 9.4(a), 9.5(a) and 9.6(a) match well the respective histogram plots in Figures 9.4(b), 9.5(b) and

$\mathbf{C}_1$	0.0531	0.0	-0.0007	-0.0004	0.0077	-0.0001
	0.0	0.0531	0.0004	-0.0007	0.0001	0.0077
	-0.0007	0.0004	0.0400	0.0	0.0018	-0.0011
	-0.0004	-0.0007	0.0	0.0400	0.0011	0.0018
	0.0077	0.0001	0.0018	0.0011	0.0353	0.0
	-0.0001	0.0077	-0.0011	0.0018	0.0	0.0353
$\mathbf{C}_2$	1.0	0.0	0.4	-0.5	0.6	-0.6
	0.0	1.0	0.5	0.4	0.6	0.6
	0.4	0.5	2.0	0.0	0.7	-0.4
	-0.5	0.4	0.0	2.0	0.4	0.7
	0.6	0.6	0.7	0.4	3.0	0.0
	-0.6	0.6	-0.4	0.7	0.0	3.0
$\mathbf{C}_3$	1.0	0.0	0.1	-0.1	0.6	-0.1
	0.0	1.0	0.1	0.1	0.1	0.6
	0.1	0.1	1.0	0.0	0.1	-0.6
	-0.1	0.1	0.0	1.0	0.6	0.1
	0.6	0.1	0.1	0.6	1.0	0.0
	-0.1	0.6	-0.6	0.1	0.0	1.0

Table 9.1 Three different covariance matrices for simulating data. Covariance matrix elements for $\mathbf{C}_1$ were estimated from an actual polarimetric SAR image and the two covariance matrices $\mathbf{C}_2$ and $\mathbf{C}_3$ were artificially chosen.

9.6(b). For the range of parameters considered in these three cases, 20 terms in the summations in Equation 7.45 and Equation 7.59 were sufficient to provide good accuracy of the PDF. More terms are needed for larger values of I_X, I_Y or I_Z, but the density falls to small values for larger values of the intensities.

The root mean square error (RMSE) between the three-dimensional histogram of the sample data and the PDF was used to numerically assess the similarity. The RMSE, $\mathcal{E}$, between the three-dimensional histogram and the PDF is given by

$$\mathcal{E} = \left[\frac{\sum_i (h_i - p_i)^2}{\sum_i p_i^2}\right]^{1/2} \tag{9.11}$$

where h_i are the binned values of the histogram, and p_i are the corresponding frequencies calculated from the PDF, i.e.

$$p_i = N\, \mathrm{f}(I_{Xi}, I_{Yi}, I_{Zi}) \Delta I_X\, \Delta I_Y\, \Delta I_Z, \tag{9.12}$$

where N is the number of samples, f is the joint density function and ΔI_X, ΔI_Y and ΔI_Z are the respective bin widths. Table 9.2 gives the RMSE values for the three cases. The small error values shown in the table are to be expected since the probability of the bin i is approximated by multiplying the PDF value at the center of the bin with the volume of the bin rather than integrating the PDF over the three dimensional volume of the bin.

	$\mathbf{C}_1$ and $n = 1$	$\mathbf{C}_2$ and $n = 3$	$\mathbf{C}_3$ and $n = 4$
$\mathcal{E}$ using actual parameters	0.0101	0.0117	0.0141

Table 9.2 The root mean square error (RMSE) (see definition in Equation 9.11) between the three-dimensional histogram and the new likelihood function for three different cases of simulated data. See Table 9.3 for the actual parameter values.

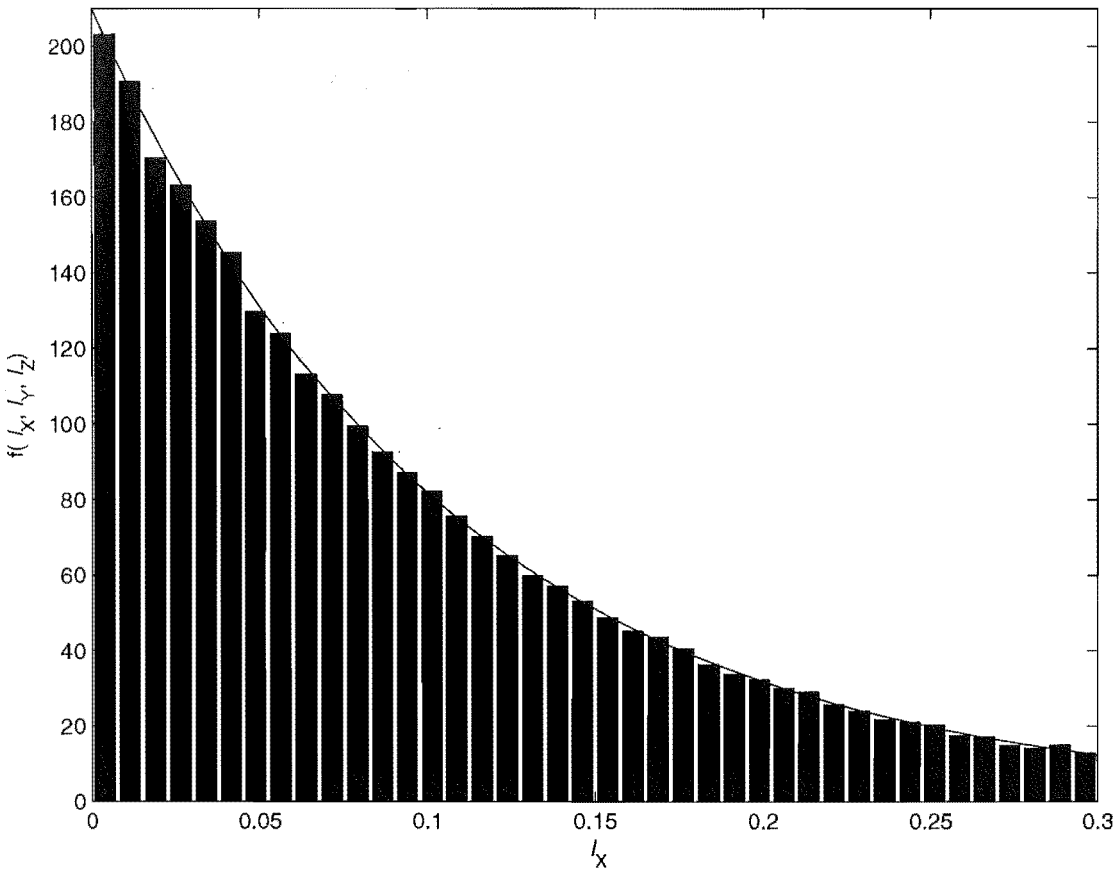


Figure 9.1 One-dimensional cut through the scaled histogram with the actual PDF overlaid for the $\mathbf{C}_1$ and $n = 1$ case with $I_Y = 0.09$ and $I_Z = 0.06$.

Parameters	$\mathbf{C}_1$ and $n = 1$		$\mathbf{C}_2$ and $n = 3$		$\mathbf{C}_3$ and $n = 4$	
	Actual	Estimate	Actual	Estimate	Actual	Estimate
a_1	19.490	19.233	1.579	1.565	1.293	1.304
b_1	25.096	25.078	0.673	0.673	1.638	1.679
c_1	29.327	29.300	0.469	0.464	1.500	1.526
d_1	0.628	0.810	0.421	0.418	0.628	0.662
d_2	1.605	1.813	0.144	0.136	0.856	0.899
d_3	4.310	3.935	0.388	0.376	0.464	0.488
d_4	4.3450	5.7758	0.0068	0.0050	0.2192	0.2685
$\lvert\mathbf{C}\rvert$	$5.23 \cdot 10^{-9}$	$5.33 \cdot 10^{-9}$	11.479	11.686	0.336	0.325

Table 9.3 Actual and estimated parameter values for three different simulations using $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ (Table 9.1). The actual parameters were calculated using Equations 7.6 - 7.9. For the estimated parameters 10^4 simulated intensity samples were used.

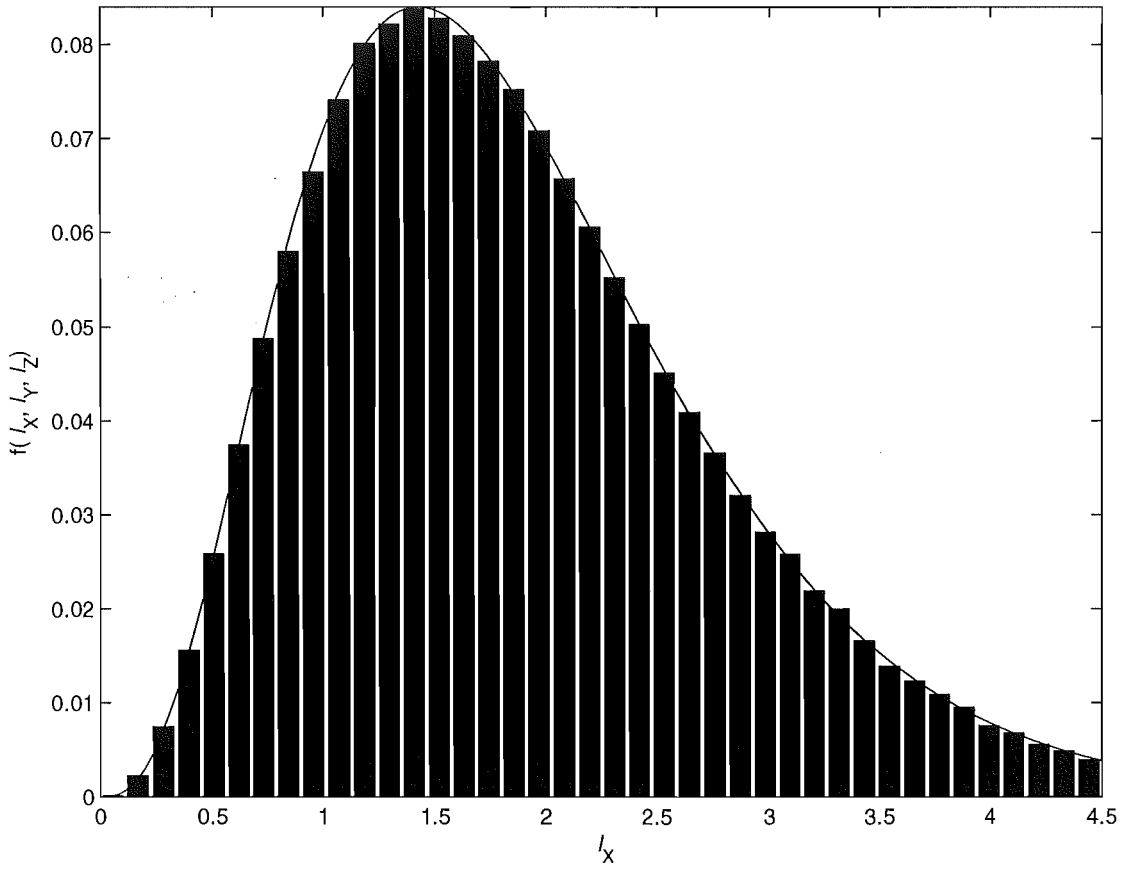


Figure 9.2 One-dimensional cut through the scaled histogram with the actual PDF overlaid for the $\mathbf{C}_2$ and $n = 3$ case with $I_Y = 1.40$ and $I_Z = 0.95$.

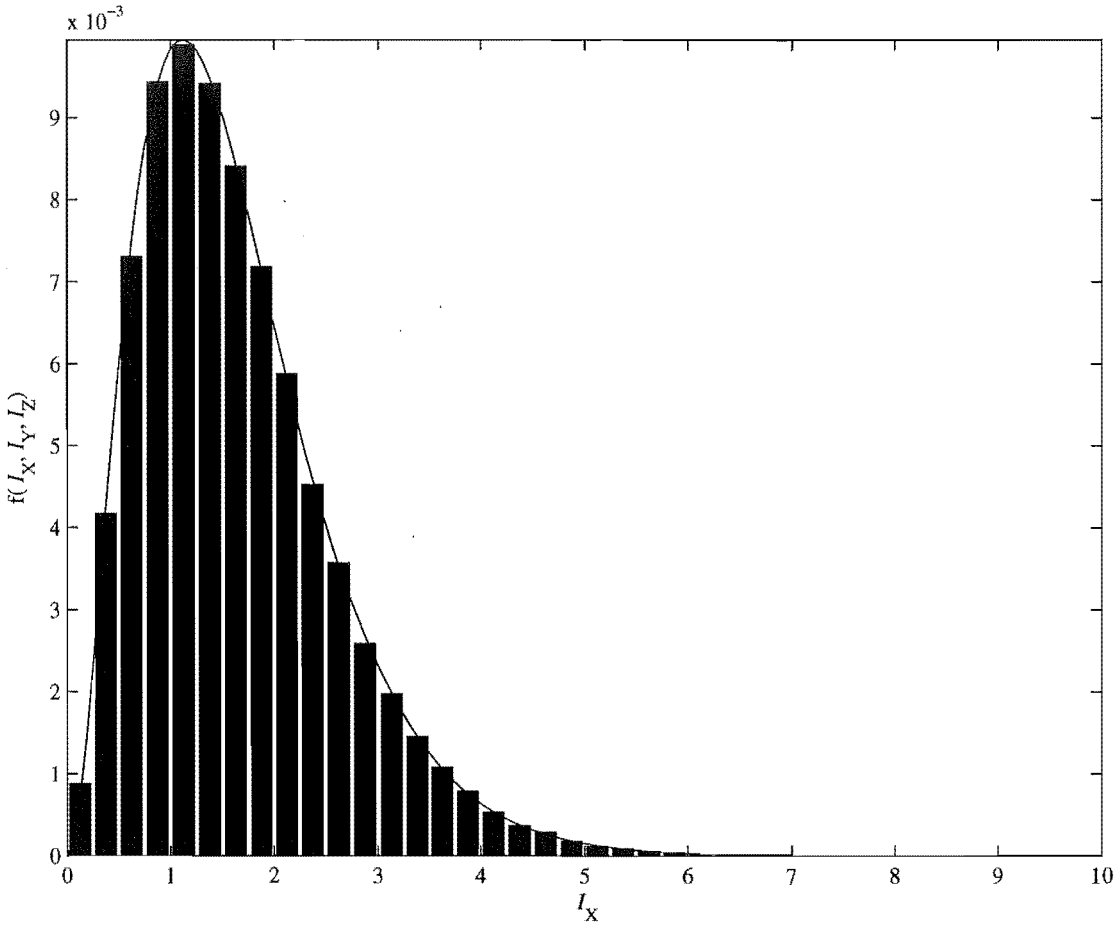
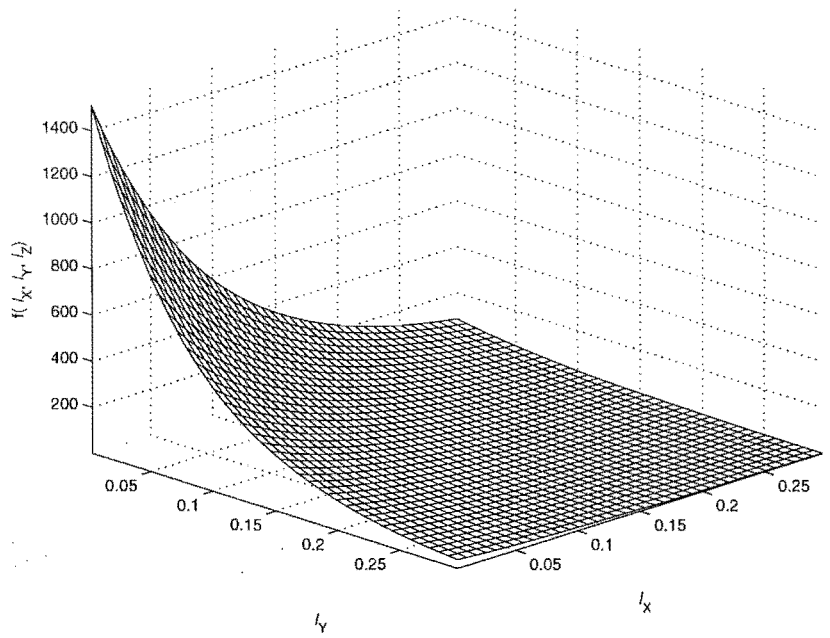
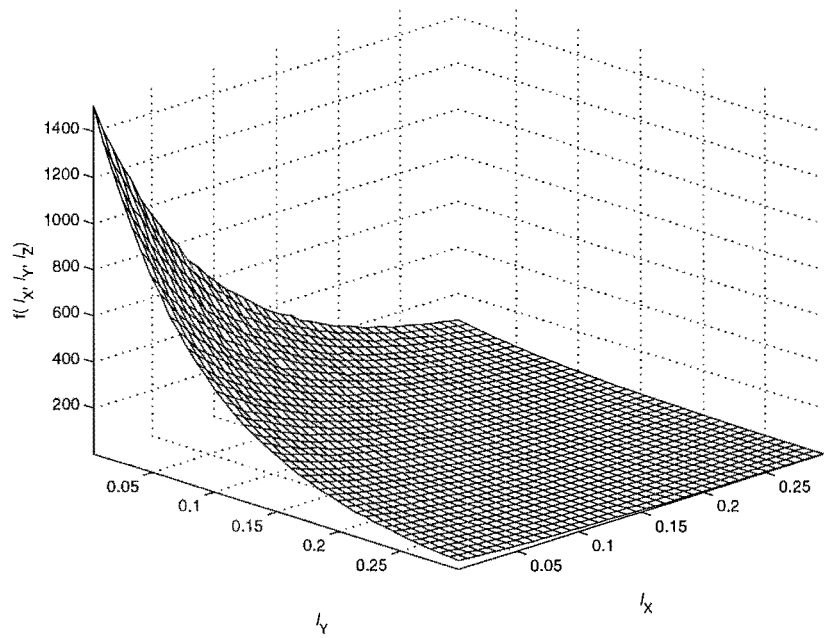


Figure 9.3 One-dimensional cut through the scaled histogram with the actual PDF overlaid for the $\mathbf{C}_3$ and $n = 4$ case with $I_Y = 3.12$ and $I_Z = 2.12$.

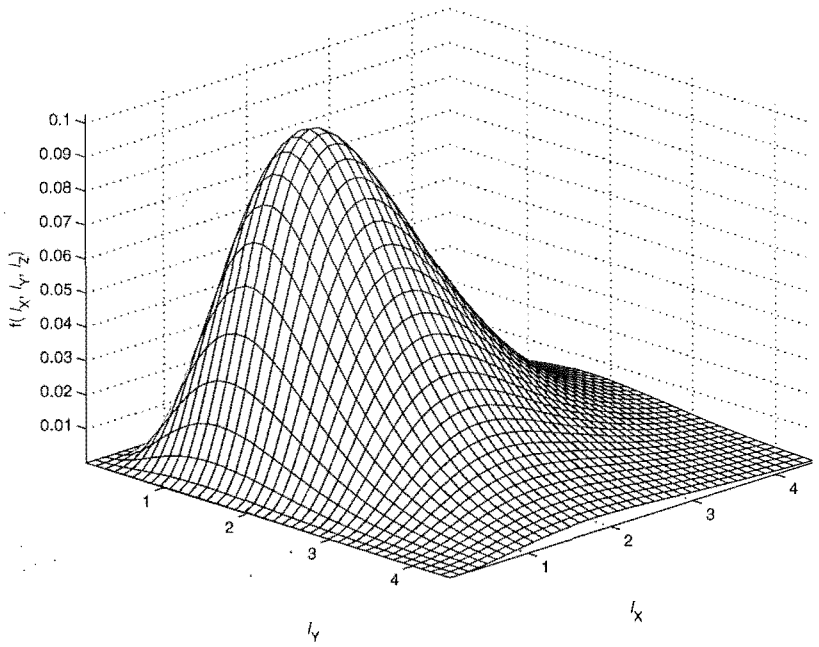


(a)

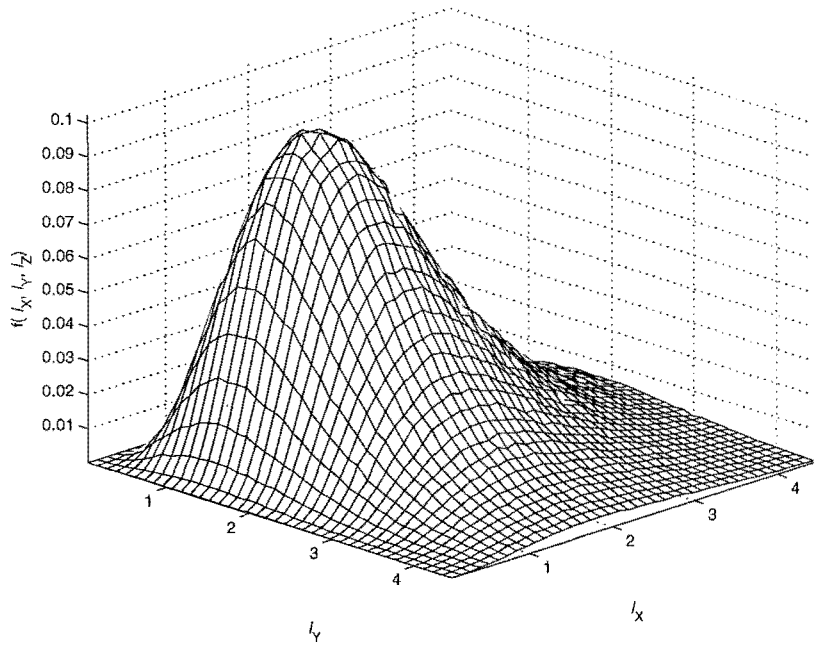


(b)

Figure 9.4 Two-dimensional plot of (a) the PDF and (b) a two-dimensional slice through the scaled histogram for the $\mathbf{C}_1$ and $n = 1$ case with $I_Z = 0.003$.

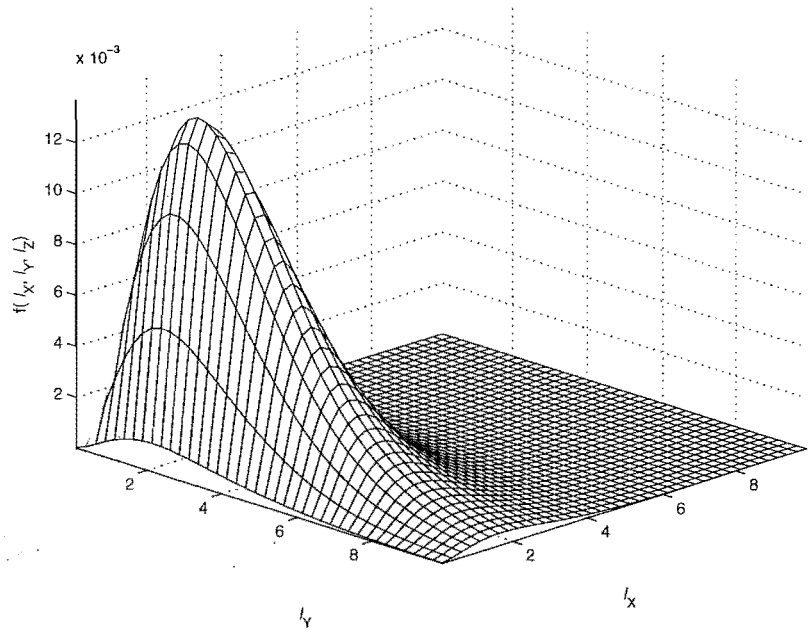


(a)

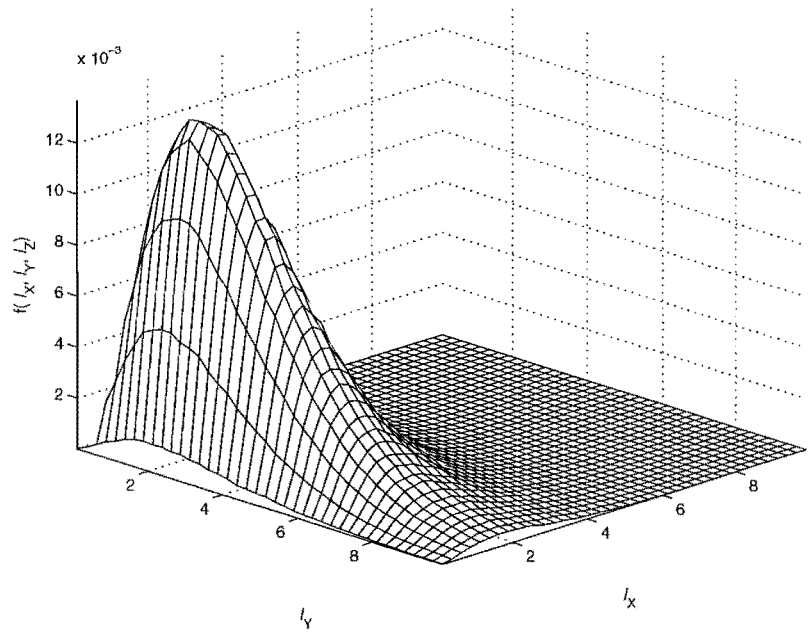


(b)

Figure 9.5 Two-dimensional plot of (a) the PDF and (b) a two-dimensional slice through the scaled histogram for the $\mathbf{C}_2$ and $n = 3$ case with $I_Z = 1.63$.



(a)



(b)

Figure 9.6 Two-dimensional plot of (a) the PDF and (b) a two-dimensional slice through the scaled histogram for the $\mathbf{C}_3$ and $n = 4$ case with $I_Z = 3.62$.

9.3 PARAMETER ESTIMATION FOR THE NEW LIKELIHOOD FUNCTION

In this section the estimation of the eight parameters for the new likelihood function is demonstrated. The method used for the parameter estimation is described in Section 7.4 on page 69. Simulated data from the three examples in the previous sections is used, therefore the actual parameter values are known and are compared to the estimates.

The likelihood function statistically describes the SAR data coming from a single homogenous type of ground cover. For example a large field would appear as a homogenous area in the SAR image. All site values belonging to that field can be seen as being samples from the same PDF with a fixed set of parameters characterising the PDF. This PDF is called the likelihood function. Given the site values from the grass field the parameters can be estimated and if the knowledge exists that this ground cover is indeed a grass field, this set of parameters can be labelled as such. Therefore in a classification task there are a number of different sets of parameters corresponding to different ground cover (estimated from a training set with prior knowledge of the ground cover), one for each class of ground cover to be classified. The parameter estimation is therefore an essential part of the application of the likelihood model in classification. Note that many types of ground cover are likely to have higher order texture components. In fact it was found in the examples given here that it was difficult to select sample data which was free of such texture components. Additional models could be incorporated for these textures, but such extensions to the model are not discussed in this thesis.

For complex SAR data the statistics are in the form of the complex Wishart distribution (see Section 4.1.4.1 on page 28) and can be characterised by the complex covariance matrix $\tilde{\mathbf{C}}$ and the number of looks n. Given only the three intensities $[I_X, I_Y, I_Z]$, corresponding to the three polarisations HH, HV and VV, the PDF describing the statistics is given in Chapter 7. This new likelihood function is characterised by the eight parameters a_1, b_1, c_1, d_1, d_2, d_3, d_4 and $|\mathbf{C}|$ and the number of looks n.

In order to show that the parameter estimation method introduced in Section 7.4 on page 69 is valid three examples with simulated data are given. SAR data is simulated in the same way as in the previous section using the same covariance matrices and number of looks as specified in Table 9.1 resulting in the parameters given in Table 9.3 indicated as 'actual'.

For the parameter estimation, 10^4 intensity samples were used and the eight parameters estimated as described in Section 7.4 on page 69. The more data samples used in the estimation the better the estimate (the lower the variance of the estimates). The sample frequency of 10^4 corresponds to a 100×100 pixel area, which is a realistically small training data set in remote sensing. Table 9.3 shows the actual parameter values evaluated from the covariance matrix $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ and the corresponding estimates for these parameter values. Note that the eight parameter values are independent of the number of looks n. Most of the estimated parameters correspond closely with the actual parameter values. The d_4 estimate is the less accurate estimate. As can be seen from the Equations in 7.106 on page 71 the d_4 parameter is more dependent on the higher order

moments of the sample data than the d_1, d_2 and d_3 parameters and a_1, b_1 and c_1 are even less influence by the higher order moments. The d_4 parameter is highly dependent on the estimated third order central moment that has a much higher variance than the lower order moments given the same number of samples. However the influence of the d_4 parameter on the new likelihood function is limited since the only dependent term is $\cos\gamma = d_4/(d_1 d_2 d_3)$, the argument of the Gegenbauer function $C_{k-n+1}^{n-1}(\cos\gamma)$ in the infinite sum term (see Equation 7.45 on page 60). Furthermore, in the first term of the sum the Gegenbauer function is of the form of $C_0^{n-1}(\cos\gamma) = 1$, independent of the $\cos\gamma$ value. A less accurate estimate of the d_4 parameter is therefore acceptable. Note however that an increased sample size will provide more accurate estimates for all parameters.

9.4 COMPARING THE NEW LIKELIHOOD FUNCTION WITH TRUE SAR DATA

In this section experimentally recorded SAR data is used to validate the new likelihood function. Data from sample areas is extracted and compared to the new likelihood function using the chi-square goodness of fit test.

There are many difficulties in validating the new likelihood function with experimental SAR data. The first difficulty lies within the SAR data. Often SAR images are oversampled and other processing of the data unknown to the end user has been performed. Then in most cases the ground truth of the imaged scene is not known and it is therefore difficult to collect a significant number of sample data from areas of approximately homogeneous ground cover. In addition the ground cover may introduce some form of texture which is not incorporated in the model and may therefore corrupt the statistics. Due to these difficulties the new likelihood function is also compared to the uncorrelated model (three uncorrelated chi-square PDFs).

9.4.1 Chi-Square goodness of fit test

Once the sample data has been collected it has to be compared to the new likelihood function in a meaningful way. Here the chi-square goodness of fit test is used. The RMSE used in the previous sections cannot be used since the sample size from the true SAR data is too small. For the chi-square test the new likelihood function needs to be integrated and the sampled data binned in a three-dimensional histogram arranging the bins in such a way that not less than 5 samples are in every bin (the chi-square test is not valid for small samples, the rule of thumb in the literature is that there should be a minimum of 5 counts per bin).

The chi-square test statistic is defined as

$$\chi^2_{\text{test}} = \sum_i \frac{(h_i - e_i)^2}{e_i} \tag{9.13}$$

where h_i is the observed frequency for bin i and e_i is the expected frequency for bin i. The expected

frequency is calculated from the new likelihood function as follows

$$e_i = N \iiint_{A_i} \mathrm{f}(I_{\mathrm{X}i}, I_{\mathrm{Y}i}, I_{\mathrm{Z}i})\, dI_{\mathrm{X}}\, dI_{\mathrm{Y}}\, dI_{\mathrm{Z}} \tag{9.14}$$

where N is the number of samples collected from the data and A_i is the volume of the three dimensional bin i. In this case all e_i are evaluated by way of approximating the integrals defined in Equation 9.14 with the Monte Carlo integration method. This is achieved easily since only samples from a six-dimensional circular Gaussian distribution have to be created (see Equation 7.10 on page 54 for the exact PDF) and then converted into intensity samples I_{X}, I_{Y} and I_{Z} as shown in Equation 7.12 on page 55. Counting all samples that lie within the volume A_i and dividing this number by the total number of samples gives an estimate of the integral in Equation 9.14. The higher the number of samples used in the Monte Carlo integration the better the estimate. Therefore

$$e_i = N \iiint_{A_i} \mathrm{f}(I_{\mathrm{X}i}, I_{\mathrm{Y}i}, I_{\mathrm{Z}i}) dI_{\mathrm{X}}\, dI_{\mathrm{Y}}\, dI_{\mathrm{Z}} \quad \approx \sum_{(I_{\mathrm{X}i}, I_{\mathrm{Y}i}, I_{\mathrm{Z}i}) \in A_i} 1. \tag{9.15}$$

The test statistics follow approximately a chi-square distribution of order $(N_{\mathrm{bins}} + 1)$ where N_{bins} is the number of bins. Therefore the hypothesis that the sample data is from a population with a specific distribution is rejected if

$$\chi^2_{\mathrm{test}} > \chi^2(\alpha_{\mathrm{sig}})_{N_{\mathrm{bins}}+1} \tag{9.16}$$

where $\chi^2(\alpha_{\mathrm{sig}})_{N_{\mathrm{bins}}+1}$ is the chi-square percentage point function of order $(N_{\mathrm{bins}} + 1)$ with a significance level α_{sig}.

9.4.2 The polarimetric SAR data set and the areas of interest

Figures 9.7 to 9.9 show the experimentally recorded SAR data used in this example. The data is SIR-C data which stands for Spaceborne Imaging Radar-C-band[1]. The C-band radar signals have a wavelength of 5.8cm. This data was supplied in the form of three complex signals one for each polarisation (HH, HV and VV). The figures show the intensity values for the three polarisations (note that the intensities have been gamma scaled to reduce the dynamic range for displaying). Since no averaging of the intensity data is performed, it is single-look data. The area imaged is south of Lake Baikal in Kudara in Eastern Siberia, Russia. The large area in the upper left of the images is forest while the darker patches are thought to be clear-cut areas.

Nine sample areas have been selected as indicated in Figure 9.10. The data from these areas has been extracted and analysed. Table 9.4 gives the number of sites for each sample area and the effective number of looks (ENL) for each polarisation of the areas (see Equation 4.8 on page 26 for

[1]Data obtained from Landcare Research, New Zealand. The SIR-C band SAR is joint U.S.- German- Italian project and part of NASA's Mission to Planet Earth. The SIR-C SAR is an imaging radar system launched aboard the NASA Space Shuttle Endeavour in 1994.

Area	Number of samples	ENL_{HH}	ENL_{HV}	ENL_{VV}
1	107786	0.910	0.837	0.872
2	22938	0.883	0.881	0.886
3	10276	0.907	0.961	0.877
4	9166	0.925	0.940	0.914
5	7438	0.684	0.890	0.711
6	47804	0.899	0.833	0.874
7	5962	0.882	0.978	0.930
8	24859	0.948	0.871	0.884
9	33088	0.934	0.855	0.922

Table 9.4 The areas of interest and some properties of them. ENL_{HH} for examples stands for effective number of looks of the HH intensity values. For the areas of interest see Figure 9.10.

the definition of the ENL value). The ENL values have been calculated from the intensities of the complex data. As can be seen in Table 9.4 the ENL values are all smaller than the true number of looks (one). This is because the non-homogeneity of the ground contributes to the variance of the intensity. The variance of the intensity is therefore not completely due to the expected variance of the speckle noise. Note that due to the relatively small number of samples in each sample area it is impractical to display or analyse the three dimensional histogram as has been presented in Section 9.2.

Furthermore, the complex SAR data of each area has been analysed with regard to the circular Gaussian behaviour. The covariance matrix $\mathbf{C} = \mathrm{E}[\mathbf{W}\,\mathbf{W}^T]$ where $\mathbf{W} = [X_\mathrm{R}, X_\mathrm{I}, Y_\mathrm{R}, Y_\mathrm{I}, Z_\mathrm{R}, Z_\mathrm{I}]^T$ has been calculated. The properties of $\mathbf{C}$, for example symmetries and local symmetries, have been described in Section 4.1.4 on page 26. Column two in Table 9.5, "Covariance matrix $\mathbf{C} = \mathrm{E}[\mathbf{W}\,\mathbf{W}^T]$", gives these covariance matrices. It is clear from the values that the circular Gaussian assumption is not valid for this complex SAR data. This is due to many factors, including: the inability to choose samples from a homogenous ground cover; and not enough scatterers or not sufficiently randomly distributed scatterers in one ground resolution cell. Furthermore the covariance matrix is an estimate and therefore bound to be inexact especially for small sample numbers.

Assuming circular Gaussian properties in the covariance matrix and calculating the complex covariance matrix $\tilde{\mathbf{C}} = \mathrm{E}[\tilde{\mathbf{W}}\,\tilde{\mathbf{W}}^\dagger]$ with $\tilde{\mathbf{W}} = [\tilde{X}, \tilde{Y}, \tilde{Z}]^T$ a covariance matrix $\mathbf{C}$ as given in Equation 4.13 on page 27 can be formed using the values from Equation 4.14. These covariance matrices are given in column three of Table 9.5 under "Closest circular Gaussian covariance matrix". This covariance matrix can be seen as the closest equivalent to the true covariance matrix displaying circular Gaussian properties. As can be seen in the table the differences between the two covariance matrices are in most cases very small.

In the next section the chi-square goodness of fit test is used to analyse the sample data with regard to the new likelihood function.

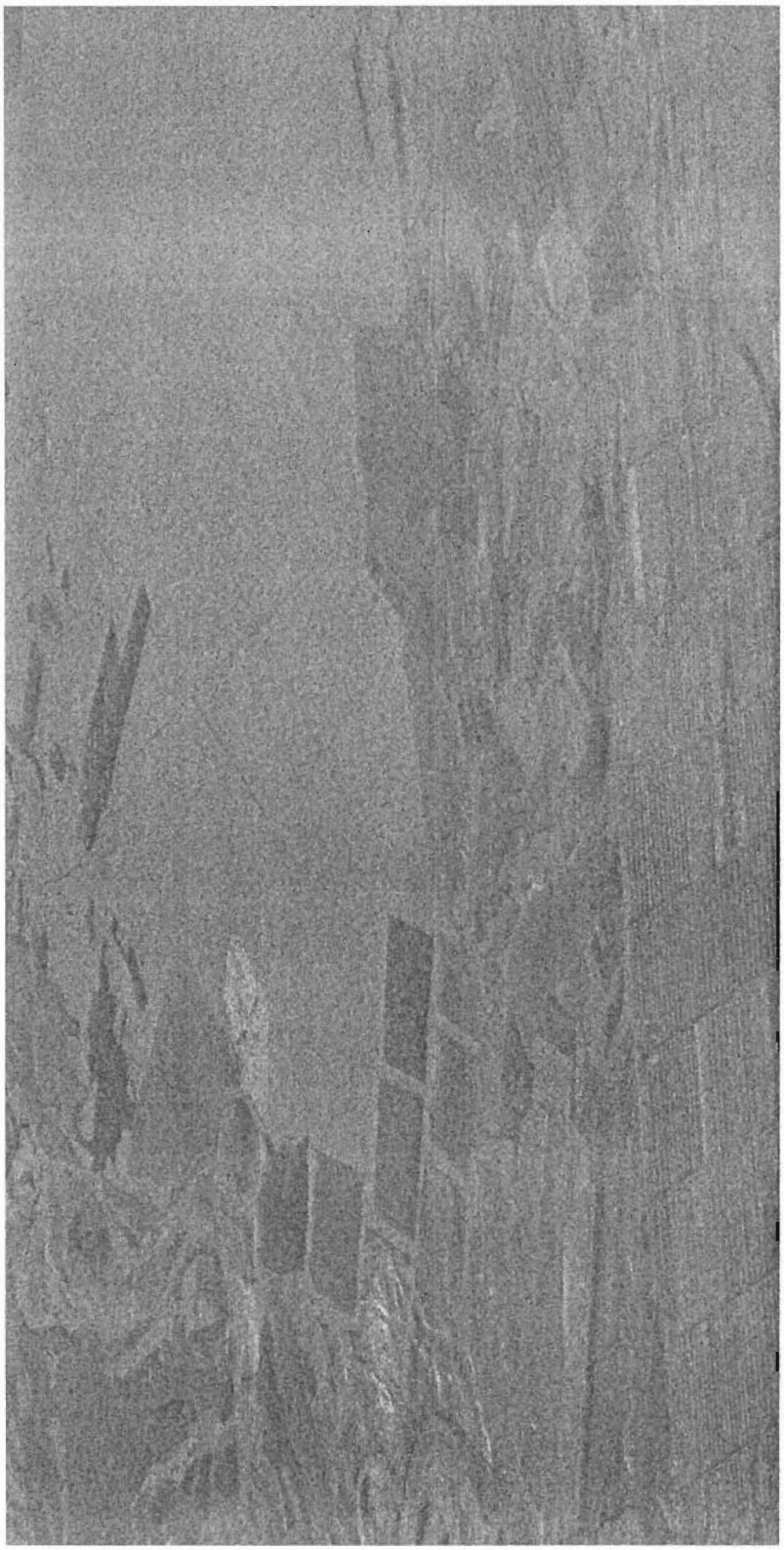


Figure 9.7 "SIR-C single look data Baikal Russia 10.07.1994" is of an area south of Lake Baikal in Kudara in Eastern Siberia, Russia. The figure shows the intensity of the HH signal. Note that a gamma scaling was used to reduce the dynamic range of the signals for display.

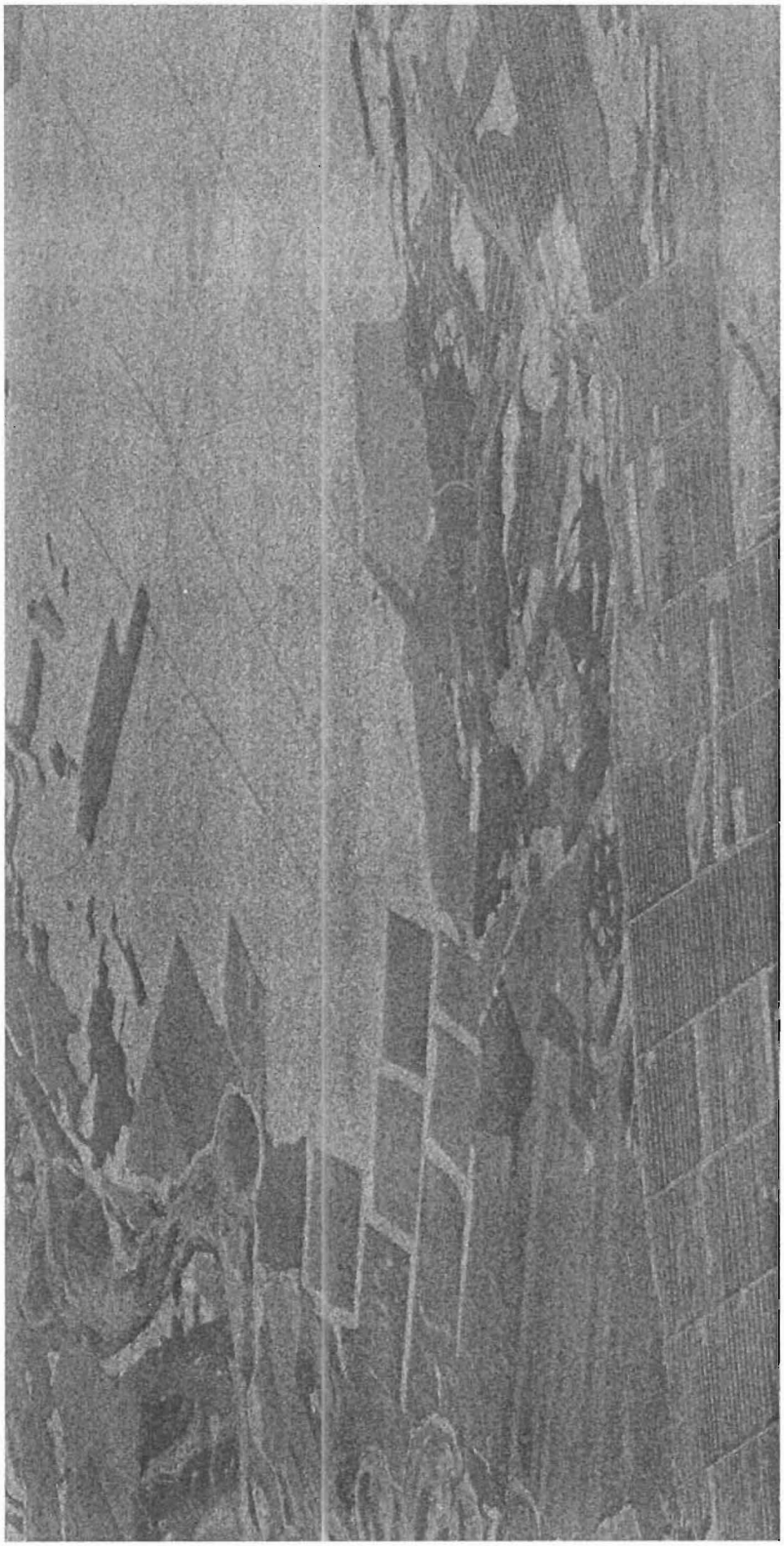


Figure 9.8 "SIR-C single look data Baikal Russia 10.07.1994" is of an area south of Lake Baikal in Kudara in Eastern Siberia, Russia. The figure shows the intensity of the HV signal. Note that a gamma scaling was used to reduce the dynamic range of the signals for display

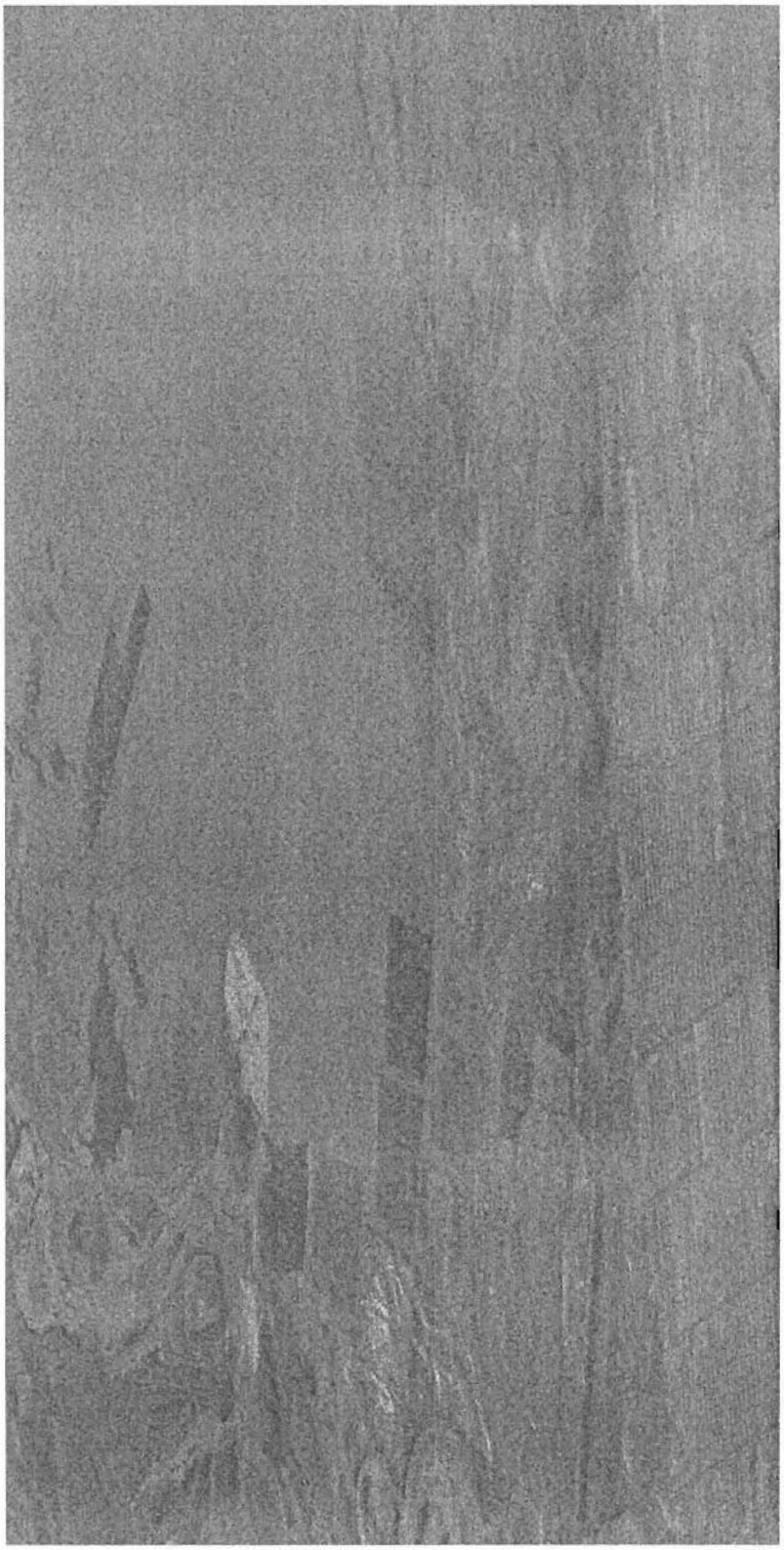


Figure 9.9 "SIR-C single look data Baikal Russia 10.07.1994" is of an area south of Lake Baikal in Kudara in Eastern Siberia, Russia. The figure shows the intensity of the VV signal. Note that a gamma scaling was used to reduce the dynamic range of the signals for display.

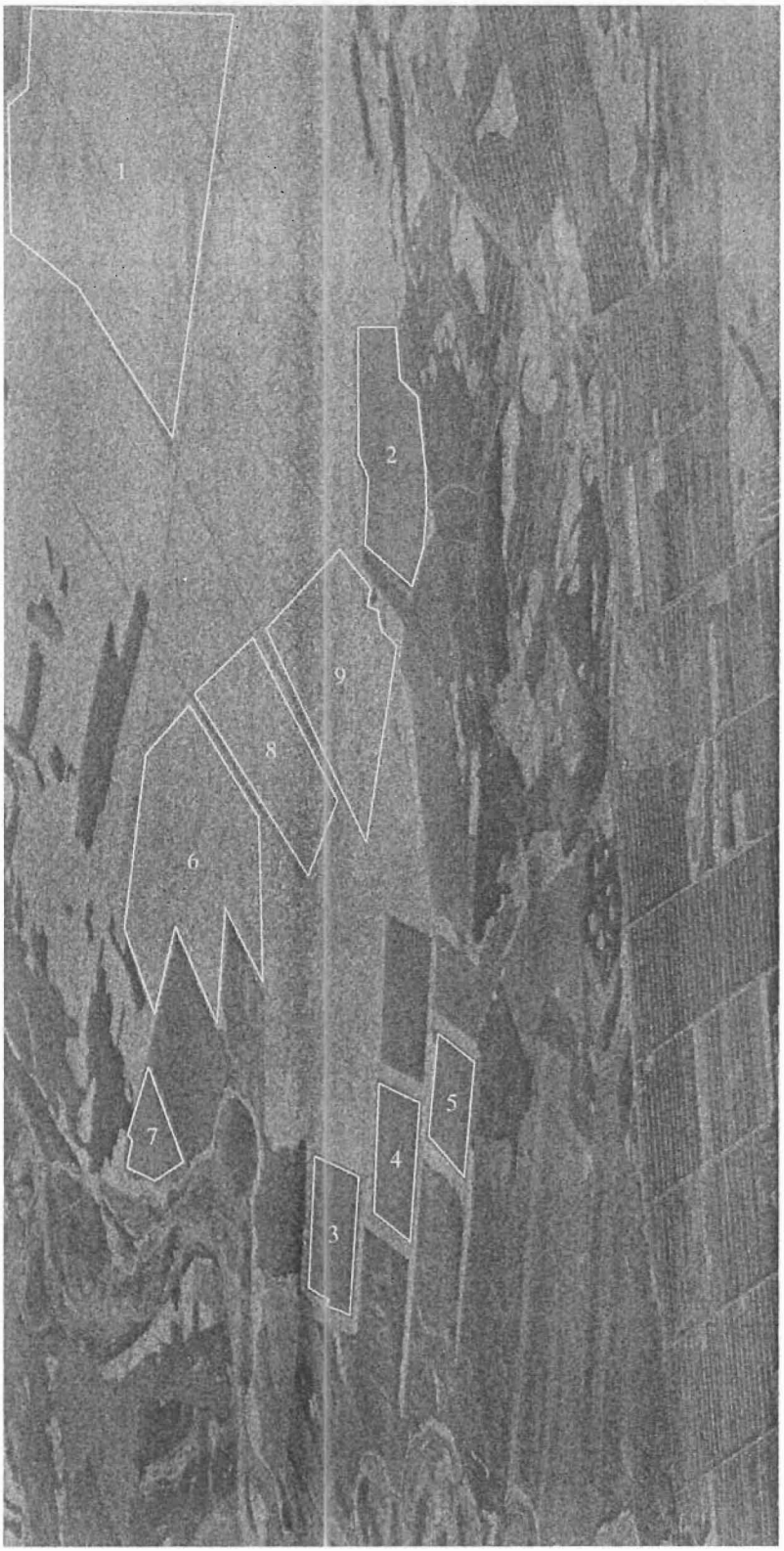


Figure 9.10 The white polygons indicate areas of interest with approximately homogenous ground cover and their respective numbers.

	Covariance matrix $\mathbf{C} = \mathrm{E}[\mathbf{W}\,\mathbf{W}^T]$						Closest circular Gaussian covariance matrix					
1	0.0745	0.0006	-0.0002	0.0006	0.0068	-0.0148	0.0742	0.0000	-0.0001	0.0005	0.0069	-0.0148
	0.0006	0.0738	-0.0003	-0.0000	0.0149	0.0070	0.0000	0.0742	-0.0005	-0.0001	0.0148	0.0069
	-0.0002	-0.0003	0.0107	0.0000	-0.0003	-0.0003	-0.0001	-0.0005	0.0107	0.0000	-0.0003	-0.0004
	0.0006	-0.0000	0.0000	0.0106	0.0005	-0.0003	0.0005	-0.0001	0.0000	0.0107	0.0004	-0.0003
	0.0068	0.0149	-0.0003	0.0005	0.0370	0.0003	0.0069	0.0148	-0.0003	0.0004	0.0369	0.0000
	-0.0148	0.0070	-0.0003	-0.0003	0.0003	0.0368	-0.0148	0.0069	-0.0004	-0.0003	0.0000	0.0369
2	0.0201	-0.0001	0.0001	0.0007	-0.0040	-0.0152	0.0203	0.0000	0.0002	0.0008	-0.0041	-0.0154
	-0.0001	0.0205	-0.0008	0.0002	0.0156	-0.0041	0.0000	0.0203	-0.0008	0.0002	0.0154	-0.0041
	0.0001	-0.0008	0.0018	0.0000	-0.0008	0.0004	0.0002	-0.0008	0.0018	0.0000	-0.0008	0.0003
	0.0007	0.0002	0.0000	0.0018	-0.0002	-0.0008	0.0008	0.0002	0.0000	0.0018	-0.0003	-0.0008
	-0.0040	0.0156	-0.0008	-0.0002	0.0208	-0.0001	-0.0041	0.0154	-0.0008	-0.0003	0.0207	0.0000
	-0.0152	-0.0041	0.0004	-0.0008	-0.0001	0.0205	-0.0154	-0.0041	0.0003	-0.0008	0.0000	0.0207
3	0.0133	0.0001	0.0001	0.0004	-0.0028	-0.0110	0.0134	0.0000	0.0001	0.0005	-0.0027	-0.0110
	0.0001	0.0135	-0.0005	0.0001	0.0110	-0.0026	0.0000	0.0134	-0.0005	0.0001	0.0110	-0.0027
	0.0001	-0.0005	0.0008	-0.0000	-0.0005	0.0001	0.0001	-0.0005	0.0008	0.0000	-0.0005	0.0001
	0.0004	0.0001	-0.0000	0.0008	-0.0001	-0.0005	0.0005	0.0001	0.0000	0.0008	-0.0001	-0.0005
	-0.0028	0.0110	-0.0005	-0.0001	0.0139	0.0002	-0.0027	0.0110	-0.0005	-0.0001	0.0139	0.0000
	-0.0110	-0.0026	0.0001	-0.0005	0.0002	0.0139	-0.0110	-0.0027	0.0001	-0.0005	0.0000	0.0139
4	0.0115	0.0003	0.0002	0.0005	-0.0023	-0.0088	0.0115	0.0000	0.0002	0.0004	-0.0024	-0.0090
	0.0003	0.0115	-0.0003	0.0002	0.0092	-0.0026	0.0000	0.0115	-0.0004	0.0002	0.0090	-0.0024
	0.0002	-0.0003	0.0008	0.0000	-0.0004	-0.0001	0.0002	-0.0004	0.0008	0.0000	-0.0004	-0.0001
	0.0005	0.0002	0.0000	0.0008	0.0001	-0.0005	0.0004	0.0002	0.0000	0.0008	0.0001	-0.0004
	-0.0023	0.0092	-0.0004	0.0001	0.0125	-0.0001	-0.0024	0.0090	-0.0004	0.0001	0.0122	0.0000
	-0.0088	-0.0026	-0.0001	-0.0005	-0.0001	0.0119	-0.0090	-0.0024	-0.0001	-0.0004	0.0000	0.0122
5	0.0166	0.0000	0.0002	0.0003	-0.0032	-0.0155	0.0168	0.0000	0.0002	0.0003	-0.0032	-0.0157
	0.0000	0.0170	-0.0003	0.0002	0.0160	-0.0032	0.0000	0.0168	-0.0003	0.0002	0.0157	-0.0032
	0.0002	-0.0003	0.0006	-0.0000	-0.0004	-0.0001	0.0002	-0.0003	0.0006	0.0000	-0.0004	-0.0001
	0.0003	0.0002	-0.0000	0.0006	0.0001	-0.0003	0.0003	0.0002	0.0000	0.0006	0.0001	-0.0004
	-0.0032	0.0160	-0.0004	0.0001	0.0201	-0.0000	-0.0032	0.0157	-0.0004	0.0001	0.0199	0.0000
	-0.0155	-0.0032	-0.0001	-0.0003	-0.0000	0.0197	-0.0157	-0.0032	-0.0001	-0.0004	0.0000	0.0199
6	0.0813	-0.0004	0.0004	0.0006	0.0093	-0.0162	0.0818	0.0000	0.0001	0.0005	0.0092	-0.0162
	-0.0004	0.0824	-0.0004	-0.0002	0.0161	0.0091	0.0000	0.0818	-0.0005	0.0001	0.0162	0.0092
	0.0004	-0.0004	0.0111	-0.0002	-0.0002	-0.0003	0.0001	-0.0005	0.0110	0.0000	-0.0002	-0.0003
	0.0006	-0.0002	-0.0002	0.0110	0.0002	-0.0003	0.0005	0.0001	0.0000	0.0110	0.0003	-0.0002
	0.0093	0.0161	-0.0002	0.0002	0.0379	0.0002	0.0092	0.0162	-0.0002	0.0003	0.0381	0.0000
	-0.0162	0.0091	-0.0003	-0.0003	0.0002	0.0383	-0.0162	0.0092	-0.0003	-0.0002	0.0000	0.0381
7	0.0151	-0.0002	-0.0002	0.0007	-0.0019	-0.0113	0.0148	0.0000	-0.0001	0.0007	-0.0015	-0.0109
	-0.0002	0.0144	-0.0006	-0.0000	0.0105	-0.0010	0.0000	0.0148	-0.0007	-0.0001	0.0109	-0.0015
	-0.0002	-0.0006	0.0010	0.0000	-0.0004	0.0001	-0.0001	-0.0007	0.0010	0.0000	-0.0005	0.0000
	0.0007	-0.0000	0.0000	0.0010	-0.0000	-0.0006	0.0007	-0.0001	0.0000	0.0010	-0.0000	-0.0005
	-0.0019	0.0105	-0.0004	-0.0000	0.0119	0.0005	-0.0015	0.0109	-0.0005	-0.0000	0.0122	0.0000
	-0.0113	-0.0010	0.0001	-0.0006	0.0005	0.0126	-0.0109	-0.0015	0.0000	-0.0005	0.0000	0.0122
8	0.0840	-0.0011	-0.0007	0.0003	0.0067	-0.0161	0.0831	0.0000	-0.0006	0.0003	0.0069	-0.0156
	-0.0011	0.0822	-0.0003	-0.0006	0.0151	0.0070	0.0000	0.0831	-0.0003	-0.0006	0.0156	0.0069
	-0.0007	-0.0003	0.0138	0.0002	-0.0005	-0.0006	-0.0006	-0.0003	0.0140	0.0000	-0.0004	-0.0006
	0.0003	-0.0006	0.0002	0.0142	0.0006	-0.0003	0.0003	-0.0006	0.0000	0.0140	0.0006	-0.0004
	0.0067	0.0151	-0.0005	0.0006	0.0440	0.0002	0.0069	0.0156	-0.0004	0.0006	0.0445	0.0000
	-0.0161	0.0070	-0.0006	-0.0003	0.0002	0.0450	-0.0156	0.0069	-0.0006	-0.0004	0.0000	0.0445
9	0.0972	-0.0001	0.0003	0.0005	0.0111	-0.0207	0.0974	0.0000	0.0001	0.0002	0.0114	-0.0201
	-0.0001	0.0976	0.0001	-0.0000	0.0195	0.0118	0.0000	0.0974	-0.0002	0.0001	0.0201	0.0114
	0.0003	0.0001	0.0134	0.0001	-0.0004	-0.0008	0.0001	-0.0002	0.0136	0.0000	-0.0002	-0.0005
	0.0005	-0.0000	0.0001	0.0138	0.0003	0.0000	0.0002	0.0001	0.0000	0.0136	0.0005	-0.0002
	0.0111	0.0195	-0.0004	0.0003	0.0456	-0.0003	0.0114	0.0201	-0.0002	0.0005	0.0461	0.0000
	-0.0207	0.0118	-0.0008	0.0000	-0.0003	0.0466	-0.0201	0.0114	-0.0005	-0.0002	0.0000	0.0461

Table 9.5 The first column gives the number of the area of interest (see Figure 9.10). The second column displays the covariance matrices $\mathbf{C} = \mathrm{E}[\mathbf{W}\,\mathbf{W}^T]$ where $\mathbf{W} = [X_\mathrm{R},\, X_\mathrm{I},\, Y_\mathrm{R}, Y_\mathrm{I}, Z_\mathrm{R}, Z_\mathrm{I}]^T$ is made up of the real and imaginary parts of the complex SAR data. Column three shows the closest circular Gaussian covariance matrices. These covariance matrices can be formed from the complex covariance matrices $\tilde{\mathbf{C}} = \mathrm{E}[\tilde{\mathbf{W}}\,\tilde{\mathbf{W}}^\dagger]$ with $\tilde{\mathbf{W}} = [\tilde{X}, \tilde{Y}, \tilde{Z}]^T$ using the values from Equation 4.14.

9.4.3 Chi-Square goodness of fit results

In this chi-square goodness of fit test, $20 \times 20 \times 20$ bins were used for the three dimensional PDF. The bins were evenly spaced over the three axes I_X, I_Y and I_Z in such a fashion that the sum of the normalised bin values was just greater than 0.85 (the sum over the complete space would be 1). This ensured that the most important parts of the density were covered by the bins.

The covariance matrices with the circular Gaussian properties from the third column in Table 9.5 were used in the Monte Carlo integration to calculated the expected frequencies e_i for the chi-square goodness of fit test (see Equation 9.13). For the Monte Carlo integration 10^8 samples were used. The e_i values converged to stable values for this number of samples. Bins of the observed frequencies h_i which contain less than 5 samples were joined to a neighbouring bin. The same bins were combined for the expected frequencies e_i.

Table 9.6 gives the chi-square values for the different sample areas. The chi-square test values for the new likelihood function are given in the second column under χ^2_{test}. These values have been calculated using Equation 9.13.

For comparison a second model was used. This model assumed no correlation between the polarisations. The covariance matrix has zeros in all off diagonal positions. The chi-square values for this model are given in the third column in Table 9.6 under $\chi^2_{\text{test no corr}}$. The corresponding chi-square value for a significance level of $\alpha_{\text{sig}} = 0.05$ is given in column four. The hypothesis is that the sample data is from a population given by the new likelihood function. From Equation 9.16 the hypothesis is rejected at a level of significance of $\alpha_{\text{sig}} = 0.05$ when $\chi^2_{\text{test}} > \chi^2_{0.05}$. Thus only for the sample areas 3, 4, 5 and 8 the hypothesis is not rejected. Some of the χ^2_{test} values are very close above the corresponding cut off value $\chi^2_{0.05}$. For all chi-square values from the uncorrelated model the hypothesis is rejected. Further note that all χ^2_{test} values are smaller than the corresponding $\chi^2_{\text{test no corr}}$ values which indicates in all cases a better fit of the sample data to the new likelihood function than to the uncorrelated model.

It has been proven mathematically in Chapter 7 and numerically confirmed in Section 9.2 that the new likelihood function describes averaged intensity data sampled from a circular Gaussian PDF. To fail the chi-square goodness of fit test can therefore have two reasons. The first one is that the sample data provided does not fit the model. The second is that a necessary assumption for the chi-square test is not fulfilled.

As far as the sample data is concerned it has already been shown that the circular Gaussian assumption is not met for the complex data but the differences are slight (see Table 9.5). Furthermore the ENL values in Table 9.4 indicate that the observed variance in the sample areas is not solely due to the speckle noise. Extra texture might be present in the data or the areas are not chosen from a homogenous ground cover. In addition to these factors it is known that this type of SAR data is oversampled which would corrupt the data further.

The usefulness of the chi-square goodness of fit test depends on many factors. In experiments with simulated sample data which should fit the new likelihood function, it was found that the number of sample values was crucial and no fixed cut off point for the minimum necessary number

Area	χ^2_{test}	$\chi^2_{\text{test no corr}}$	$\chi^2_{0.05}$
1	4624.13	5432.09	3619.37
2	1115.50	4512.86	1093.34
3	457.50	2153.59	469.84
4	419.75	1463.62	456.10
5	336.26	2401.29	345.66
6	2326.89	2656.24	2100.02
7	278.43	1001.44	264.22
8	1231.35	1300.61	1274.47
9	1857.49	2074.84	1616.96

Table 9.6 For each sample area (see Figure 9.10) the chi-square test value as described in Equation 9.13 is given. χ^2_{test} is the value with the new likelihood function as a model and $\chi^2_{\text{test no corr}}$ is the value using the model with no correlations between the polarisation. The corresponding chi-square value for a significance level of $\alpha_{\text{sig}} = 0.05$ is given under $\chi^2_{0.05}$. Note that for $\chi^2_{\text{test}} > \chi^2_{0.05}$ the hypothesis that the sample data is from a population with the specified distribution is rejected.

of samples could be given. However for some sample areas, the number of samples is low and therefore the result of the chi-square test is less conclusive. Note however that comparing the chi-square test values for the two different models is valid and indicates a better fit for the new likelihood function over the uncorrelated model as expected.

9.5 CONCLUSIONS

In this chapter the validity of the new likelihood function has been shown. In the first section it was mathematically proven that the infinite sum in the new likelihood function converges. This confirms the experimental results that the sum converges after only a small number of terms. This number of terms needed is typically smaller than 20.

It was proven in Chapter 7 that the new likelihood function describes the averaged intensity samples from a three dimensional complex circular Gaussian zero-mean density. This was confirmed in Section 9.2 where simulated sample data was compared to the new likelihood function using RMSE values for different covariance matrices and different numbers of looks. The eight parameters of the new likelihood function, namely a_1, b_1, c_1, d_1, d_2, d_3, d_4 and $|\mathbf{C}|$, can be directly calculated from the covariance matrices used in producing the simulated data. In Section 9.3 these parameters were compared to parameters estimated from a small set of sample data. 10^4 intensity samples were used in the parameter estimation. This is equivalent to a 100×100 sample area which can be considered as a realistic small training set. Most of the estimated parameters corresponded closely with the actual parameter values. However each of the estimated parameters depended in a varying degrees on higher order moments of the intensity sample data which was reflected in the variance of the estimates. The d_4 estimate was the less accurate estimate. It was however shown that the influence of the d_4 parameter on the new likelihood function is smaller than the other parameters and therefore a less accurate estimate for this parameter was acceptable. Note that an increased sample size will provide more accurate estimates for all parameters.

In the next section, Section 9.4, the new likelihood function was compared to true SAR data. The comparison of the new likelihood function and the data from the selected sample areas was done by way of the chi-square goodness of fit test. Due to the relatively small number of intensity samples the RMSE could not be used for comparison. The hypothesis tested with the chi-square test was that the sample data is from a population belonging to the new likelihood function. For 5 out of the 9 selected sample areas, this hypothesis was rejected. This was done at a level of significance of $\alpha_{\text{sig}} = 0.05$.

Several reasons for failing the chi-square test have been identified. The first one was choosing sufficient sample data from areas with homogenous ground covers. For this experiment no suitable polarimetric SAR data was available with information about the ground truth. Therefore sample areas had to be selected by visual impression of the areas which might have a homogenous ground cover. Analysis of the sample areas showed that the ENL values are below the expected number of looks $n = 1$ and that the covariance matrices of the complex SAR data did not quite show circular Gaussian behaviour. Besides the problem of selecting homogenous areas other factors might have contributed to this, such as texture in the scene (adding variance to the signal which is not part of the speckle noise), over-sampling and other processing of the SAR data unknown to the end user. Additionally the small sample size effects the validity of the chi-square test in some cases.

For comparison the same chi-square test was performed on a second model. This model is the same as the new likelihood function but assumes independency (i.e., no correlation) between all three polarisations. All of the hypotheses were rejected and the chi-square test statistic values were greater than the ones from the new likelihood function. This clearly indicates that the new likelihood function fits the sample data better than the second model (i.e., the new likelihood function assuming independency between the polarisations).

The new likelihood function describes the true and uncorrupted speckle noise process of multilook intensity data. For true SAR data, models have to be incorporated adding the effects of scene texture and other processes that effect the statistical nature of the signal. The new likelihood function can be used to separate the effect of speckle noise from other statistical processes present in the data and therefore give a better understanding of the underlying effects.

Chapter 10

PARAMETER ESTIMATION FOR THE GRF MODEL

In this Chapter the novel method of parameter estimation for GRF models, which was introduced in Chapter 8, is tested and compared to Besag's parameter estimation method. In Section 10.1 GRF realisations with known parameter values are used to form test data. The four- and the eight-neighbourhood system and different clique configurations are used. For comparison Besag's parameter estimation method is also implemented. The performance of the two parameter estimation methods is assessed. In Section 10.2 three label images are used to form test data for the comparison. Two are hand drawn two-class (binary) label images and the third one is a five-class label image generated from a Geographic Information Systems (GIS) data set of forestry blocks in Kaingaroa New Zealand. The GRF model parameters are estimated using the new parameter estimation method and Besag's parameter estimation method. In order to assess the quality of the estimated parameter sets a MAP classification, using the estimated parameters in the GRF prior model, of noisy data versions of the test label images is performed and the results measured against the true test label images. In Section 10.3 conclusions are given.

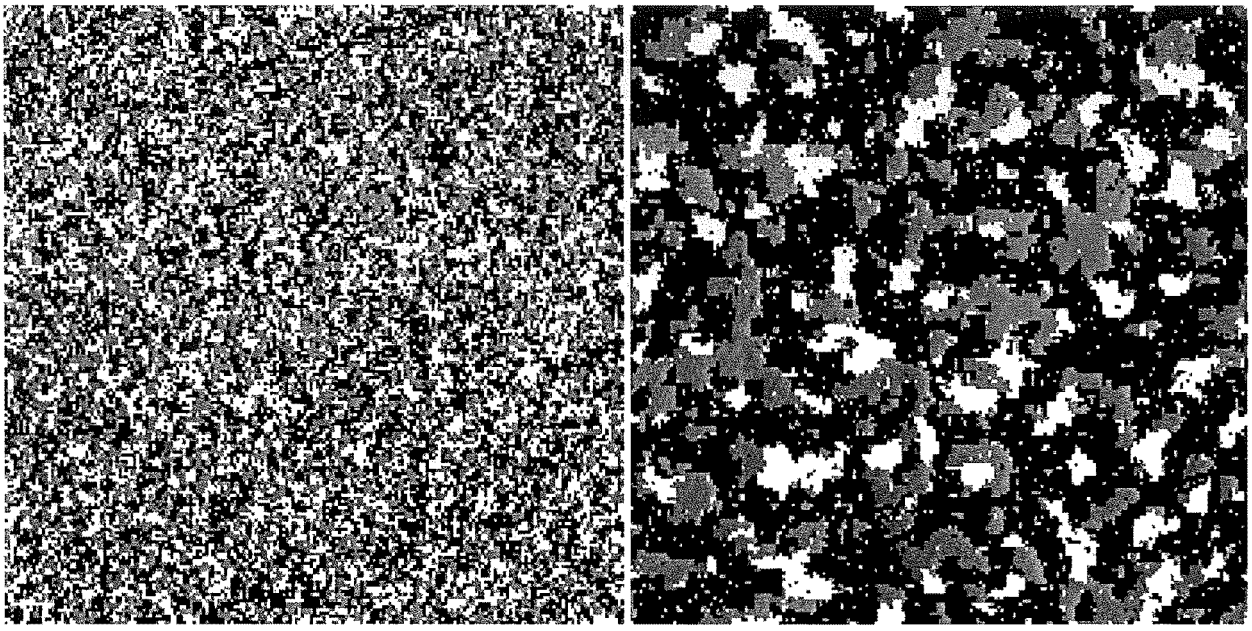
10.1 PARAMETER ESTIMATION FROM GRF REALISATIONS

The main problem in validating the new parameter estimation method is the fact that the underlying true parameters of the data are not known in advance. The data in fact might not be well described by the GRF model and therefore no perfect parameter set exists. In this section GRF realisations are used to validate the parameter estimation and additionally compare it to Besag's parameter estimation method (the coding method) described in Section 5.11.1. For the GRF realisations the parameters are known in advance and therefore the parameter estimation can be validated.

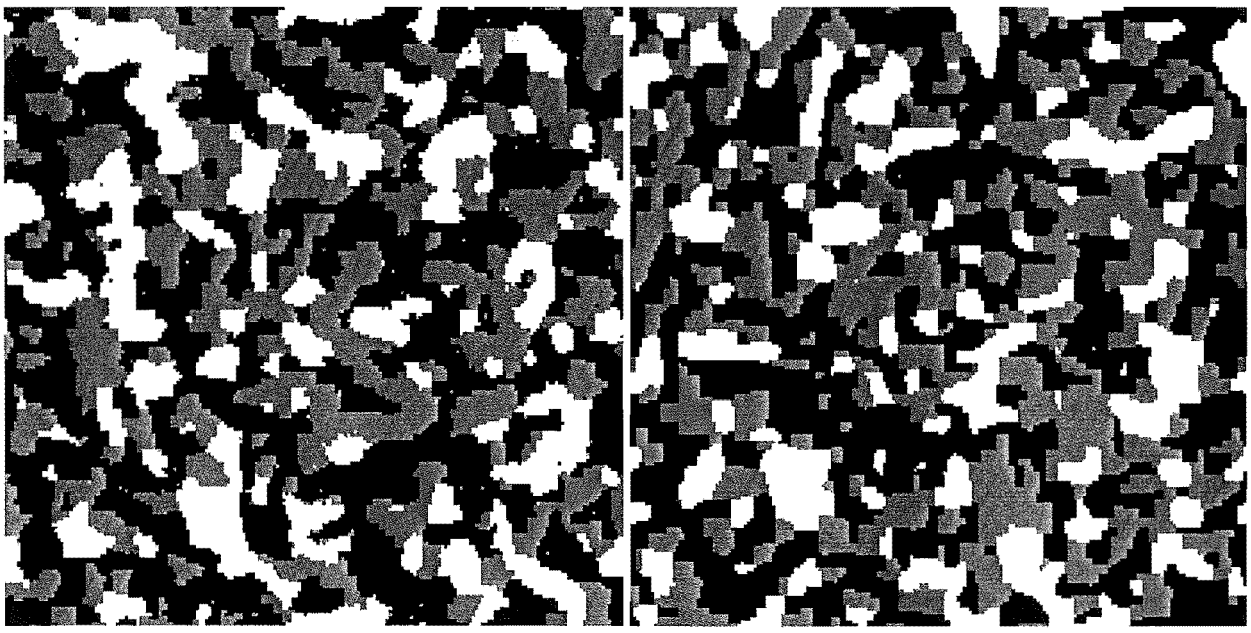
10.1.1 The GRF realisations

The Metropolis sampler is used to create realisations of the GRF for given neighbourhood systems, numbers of classes and parameter vectors. See Section 5.10.1 for details on the Metropolis sampler. Figure 10.1 through to Figure 10.7 show realisations of the GRF for different model

configurations and parameter sets. All have the size of 200 × 200 pixels. Fifty iterations (complete sweeps of the image) of the Metropolis sampler were used. After fifty iterations the changes from iteration to iteration in the image had settled to small values and experiments showed that additional iterations would not produce significantly different results in the parameter estimation. Note that the boundaries were wrapped around to the opposite boundary during the operation of the Metropolis sampler in order to avoid boundary effects. Figure 10.1 through to Figure 10.3 have the four-neighbourhood and Figure 10.4 through to Figure 10.7 the eight-neighbourhood as a basis. The corresponding parameter vector Θ is given below each image. Some of the GRF realisations have three and some four classes; the number of classes is always indicated in the captions of the figures (each class is represented by a different gray value). Table 10.1 lists all configurations of neighbourhood systems, number of classes and parameter vectors used here.

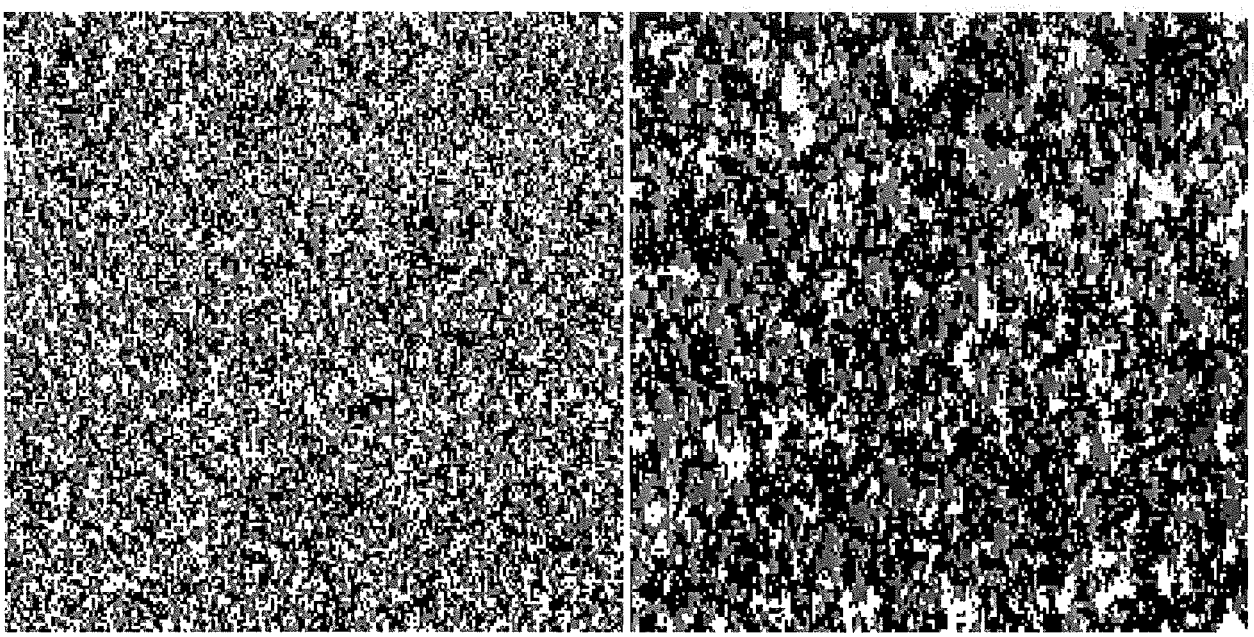


(a) $\Theta = [0.3 \ 0.3]$ (b) $\Theta = [0.6 \ 0.6]$

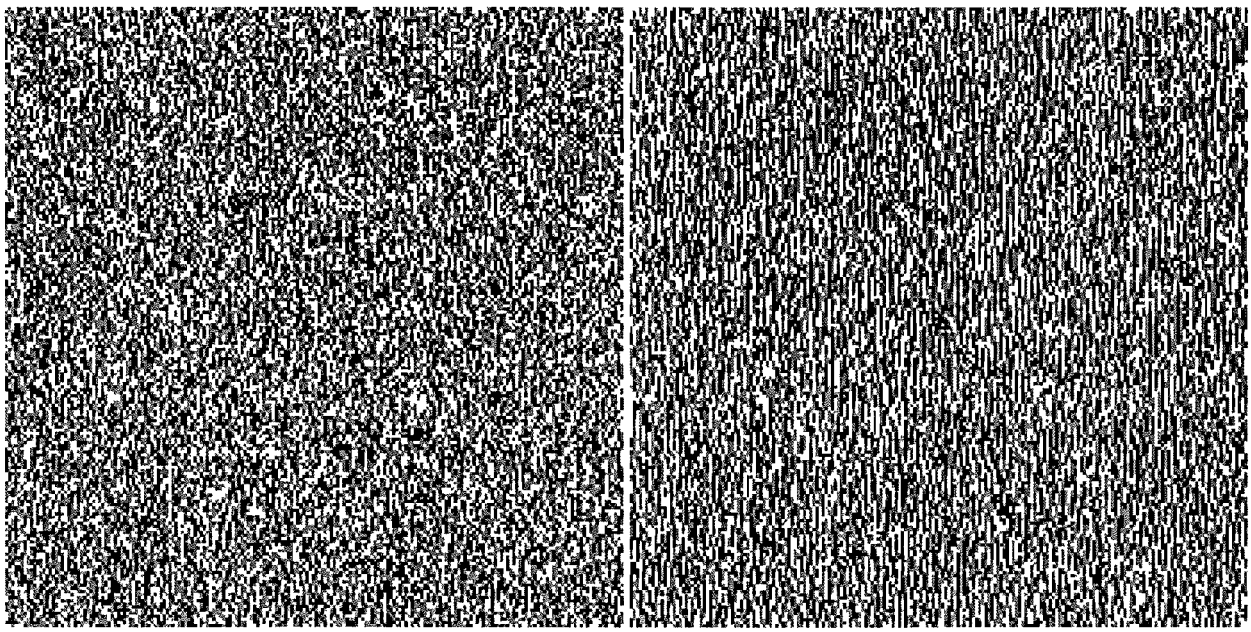


(c) $\Theta = [1.0 \ 1.0]$ (d) $\Theta = [2.0 \ 2.0]$

Figure 10.1 Realisations of a GRF. Images are of the size 200×200 pixels and contain 3 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the four-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [0.3 \ 0.3]$, (b) $\Theta = [0.6 \ 0.6]$, (c) $\Theta = [1.0 \ 1.0]$ and (d) $\Theta = [2.0 \ 2.0]$. The parameter vector is given by $\Theta = [\,|\ -\,]$ where the lines indicate the clique type (| for example is the vertical clique). Note that in this and the following figures, the boundaries were wrapped around to the opposite boundary during the operation of the Metropolis sampler in order to avoid boundary effects.

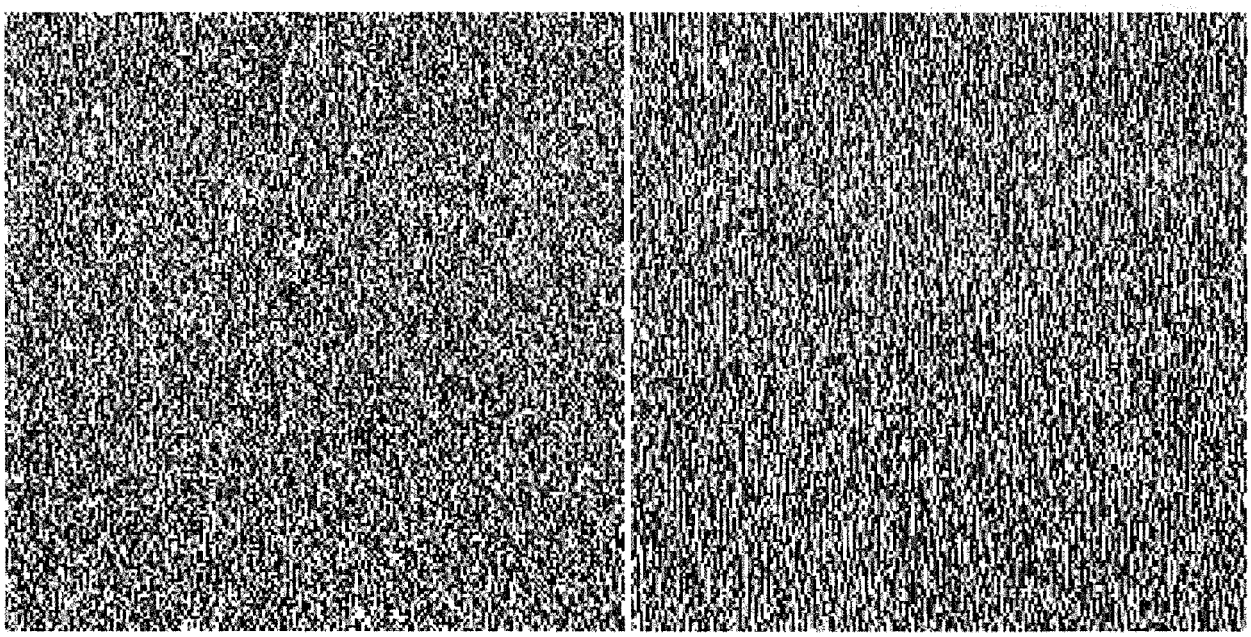


(a) $\Theta = [0.3 \;\; 0.15]$ (b) $\Theta = [0.6 \;\; 0.3]$

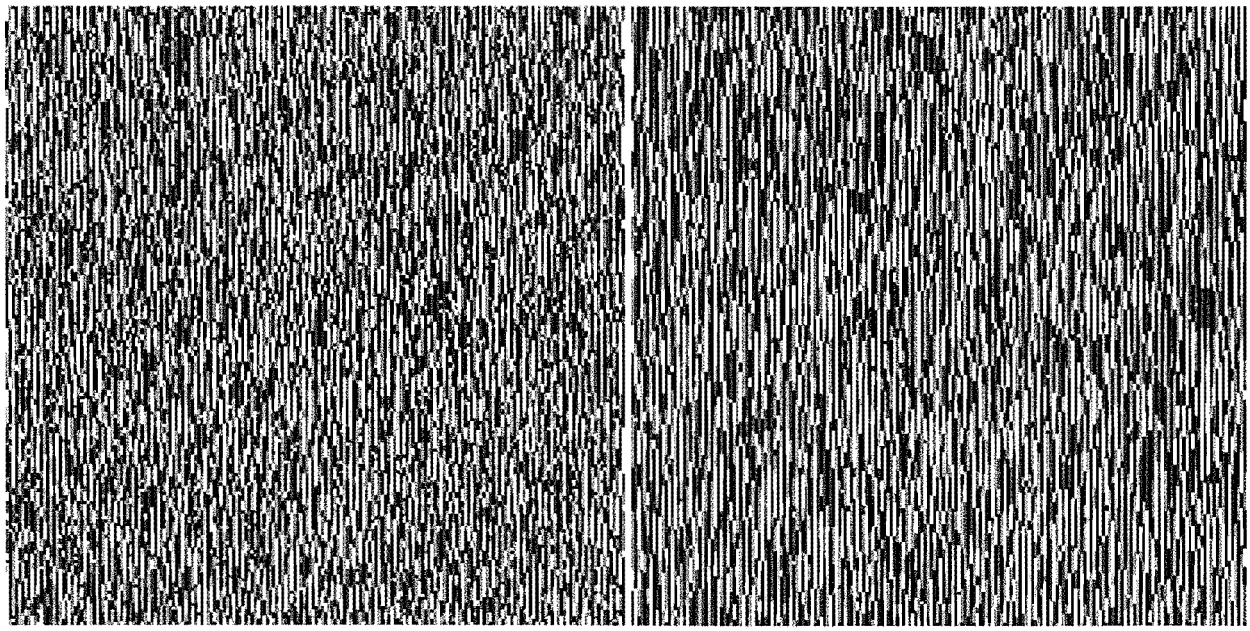


(c) $\Theta = [0.3 \;\; -0.15]$ (d) $\Theta = [0.6 \;\; -0.3]$

Figure 10.2 Realisations of a GRF. Images are of the size 200×200 pixels and contain 3 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the four-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [0.3 \;\; 0.15]$, (b) $\Theta = [0.6 \;\; 0.3]$, (c) $\Theta = [0.3 \;\; -0.15]$ and (d) $\Theta = [0.6 \;\; -0.3]$. The parameter vector is given by $\Theta = [\,|\;\; -\,]$ where the lines indicate the clique type (| for example is the vertical clique).

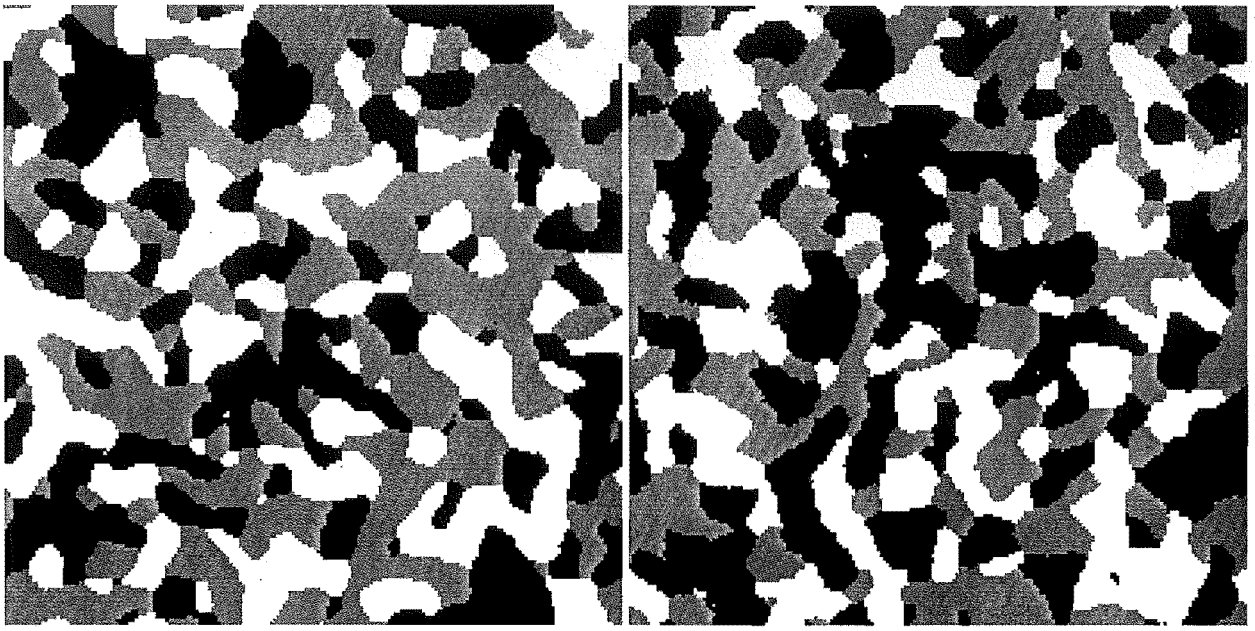


(a) $\Theta = [0.3 \quad -0.3]$ (b) $\Theta = [0.6 \quad -0.6]$

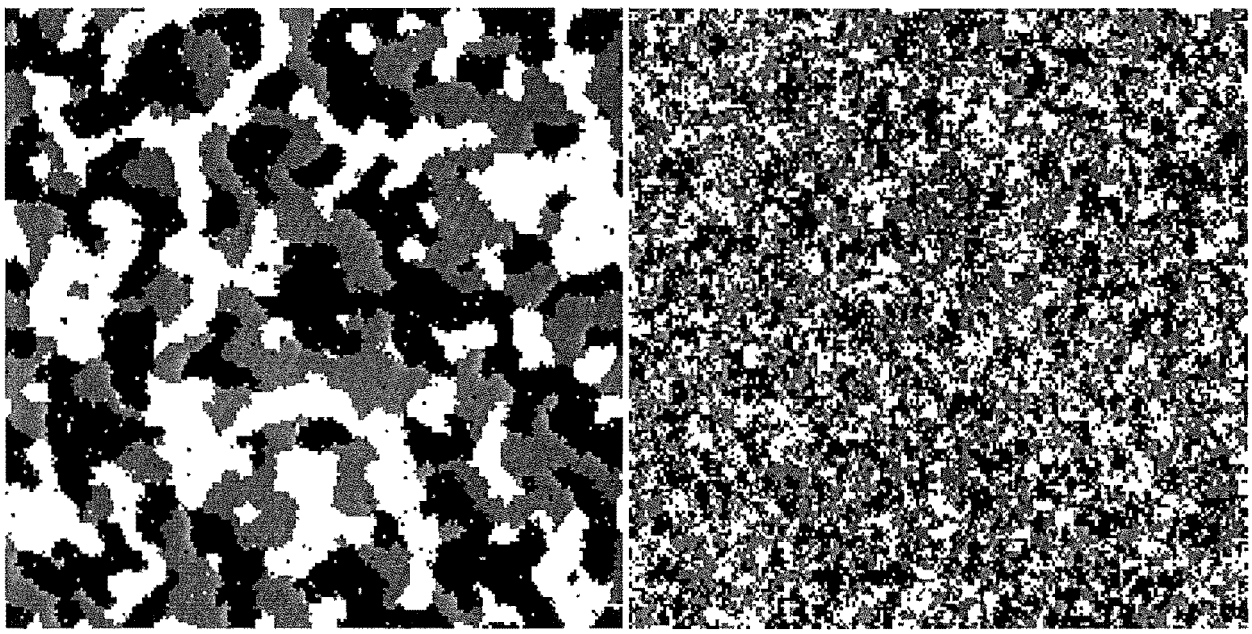


(c) $\Theta = [1 \quad -1]$ (d) $\Theta = [2 \quad -2]$

Figure 10.3 Realisations of a GRF. Images are of the size 200×200 pixels and contain 4 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the four-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [0.3 \quad -0.3]$, (b) $\Theta = [0.6 \quad -0.6]$, (c) $\Theta = [1 \quad -1]$ and (d) $\Theta = [2 \quad -2]$. The parameter vector is given by $\Theta = [\,|\quad -\,]$ where the lines indicate the clique type (| for example is the vertical clique).

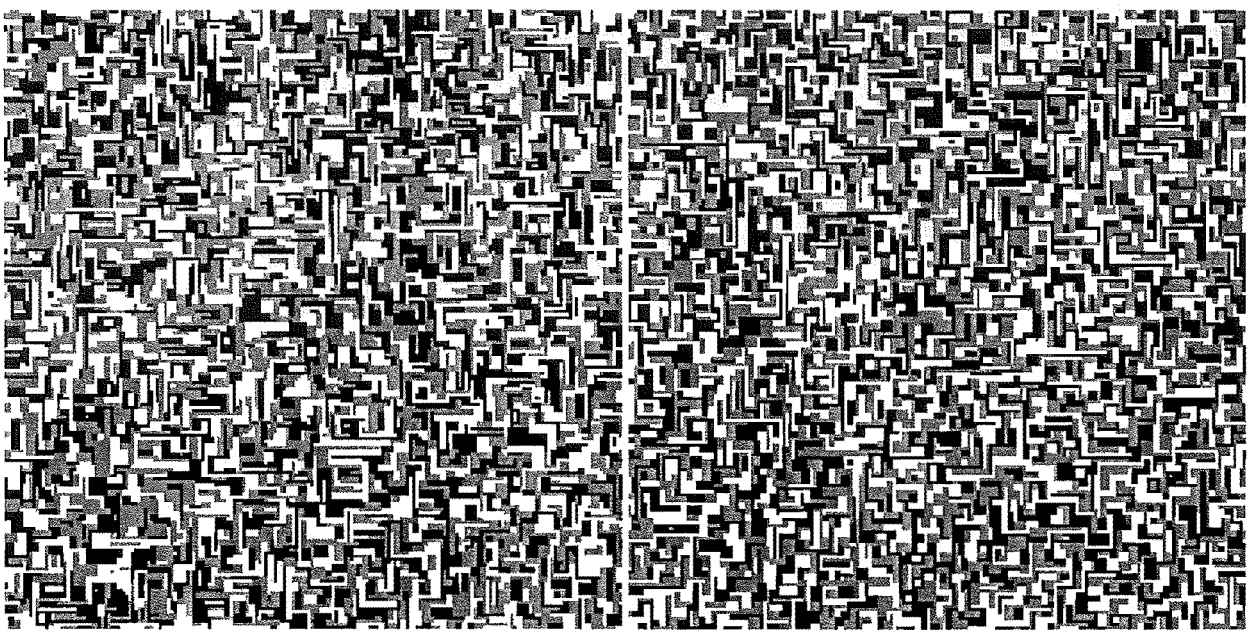


(a) $\Theta = [2\ 2\ 1\ 1]$ (b) $\Theta = [1\ 1\ 0.5\ 0.5]$

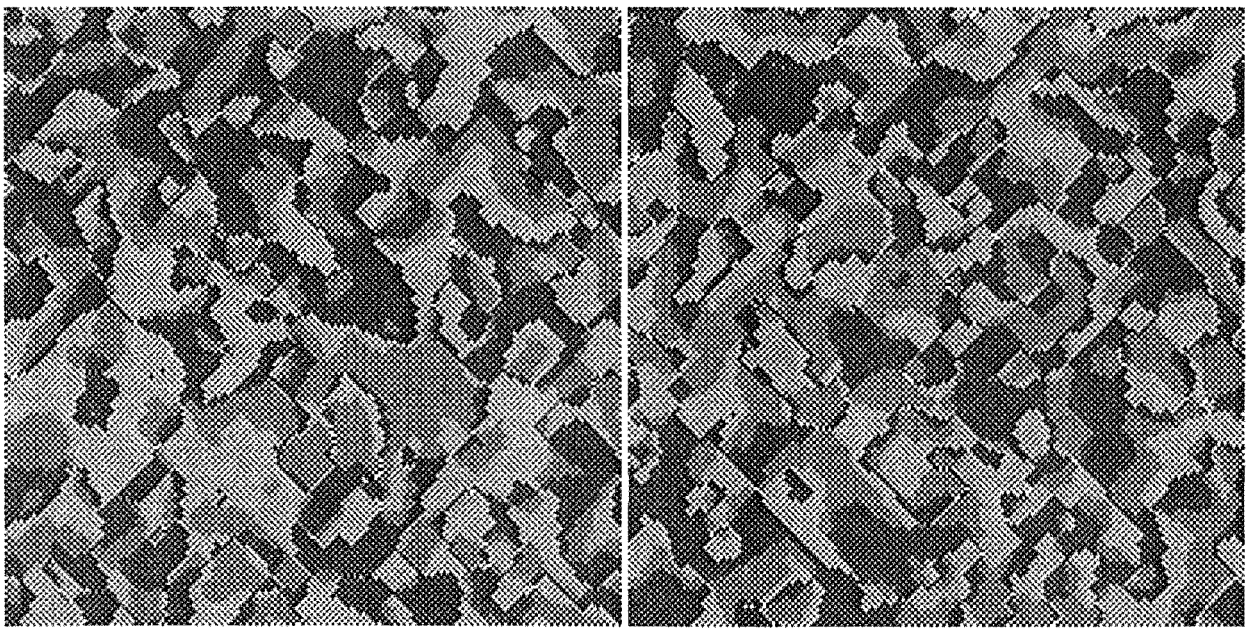


(c) $\Theta = [0.5\ 0.5\ 0.25\ 0.25]$ (d) $\Theta = [0.25\ 0.25\ 0.12\ 0.12]$

Figure 10.4 Realisations of a GRF. Images are of the size 200 × 200 pixels and contain 3 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the eight-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [2\ 2\ 1\ 1]$, (b) $\Theta = [1\ 1\ 0.5\ 0.5]$, (c) $\Theta = [0.5\ 0.5\ 0.25\ 0.25]$ and (d) $\Theta = [0.25\ 0.25\ 0.12\ 0.12]$. The parameter vector is given by $\Theta = [\,|\ -\ \backslash\ /\,]$ where the lines indicate the clique type (| for example is the vertical clique).

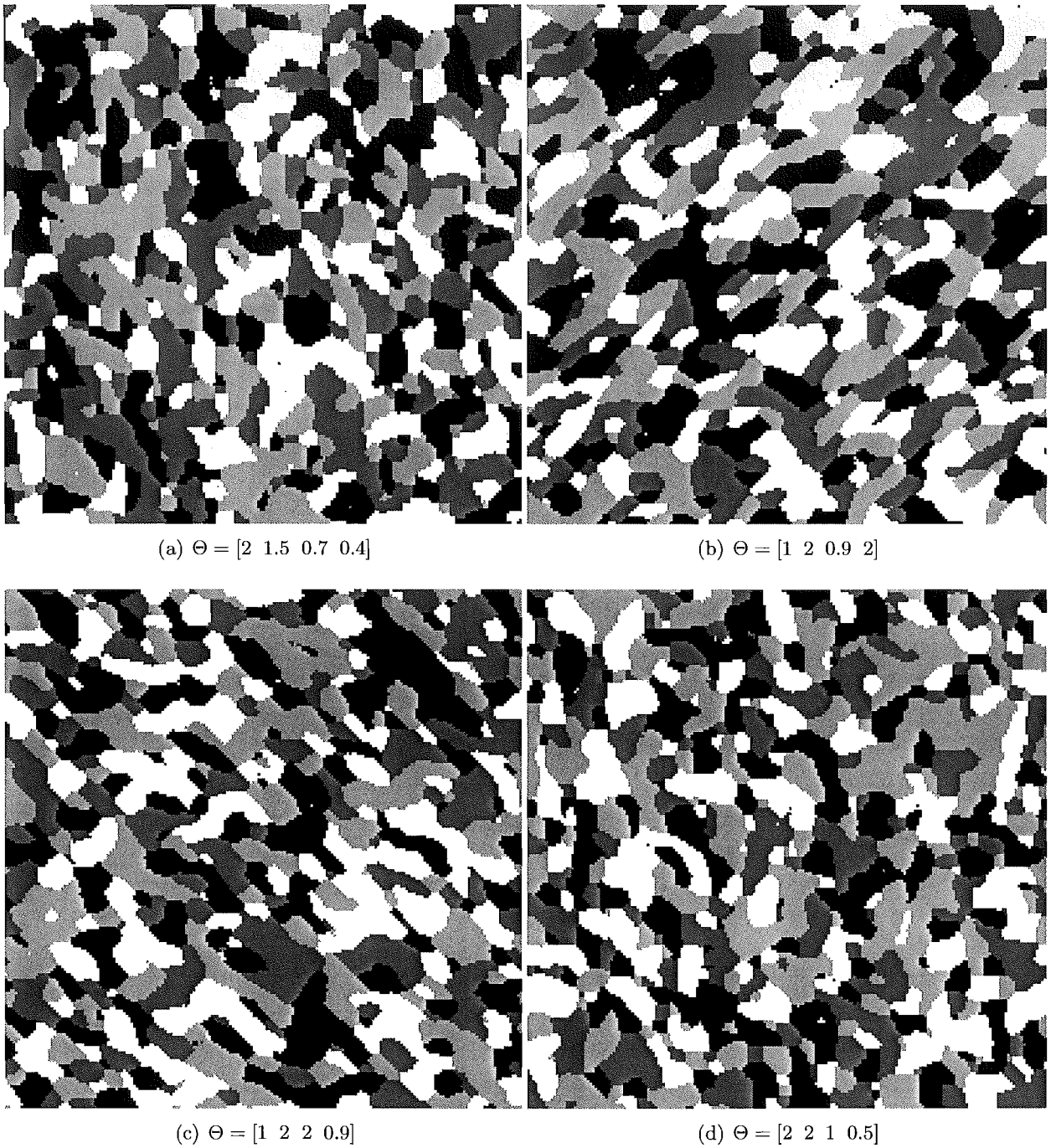


(a) $\Theta = [2\ \ 2\ \ -1\ \ -1]$ (b) $\Theta = [4\ \ 4\ \ -2\ \ -2]$



(c) $\Theta = [-1\ \ -1\ \ 2\ \ 2]$ (d) $\Theta = [-2\ \ -2\ \ 4\ \ 4]$

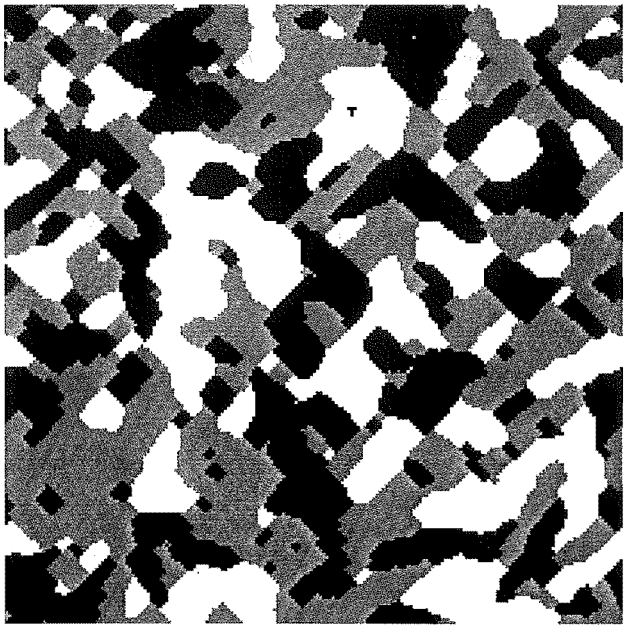
Figure 10.5 Realisations of a GRF. Images are of the size 200 × 200 pixels and contain 3 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the eight-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [2\ \ 2\ \ -1\ \ -1]$, (b) $\Theta = [4\ \ 4\ \ -2\ \ -2]$, (c) $\Theta = [-1\ \ -1\ \ 2\ \ 2]$ and (d) $\Theta = [-2\ \ -2\ \ 4\ \ 4]$. The parameter vector is given by $\Theta = [\,|\ \ -\ \ \backslash\ \ /\,]$ where the lines indicate the clique type (| for example is the vertical clique).



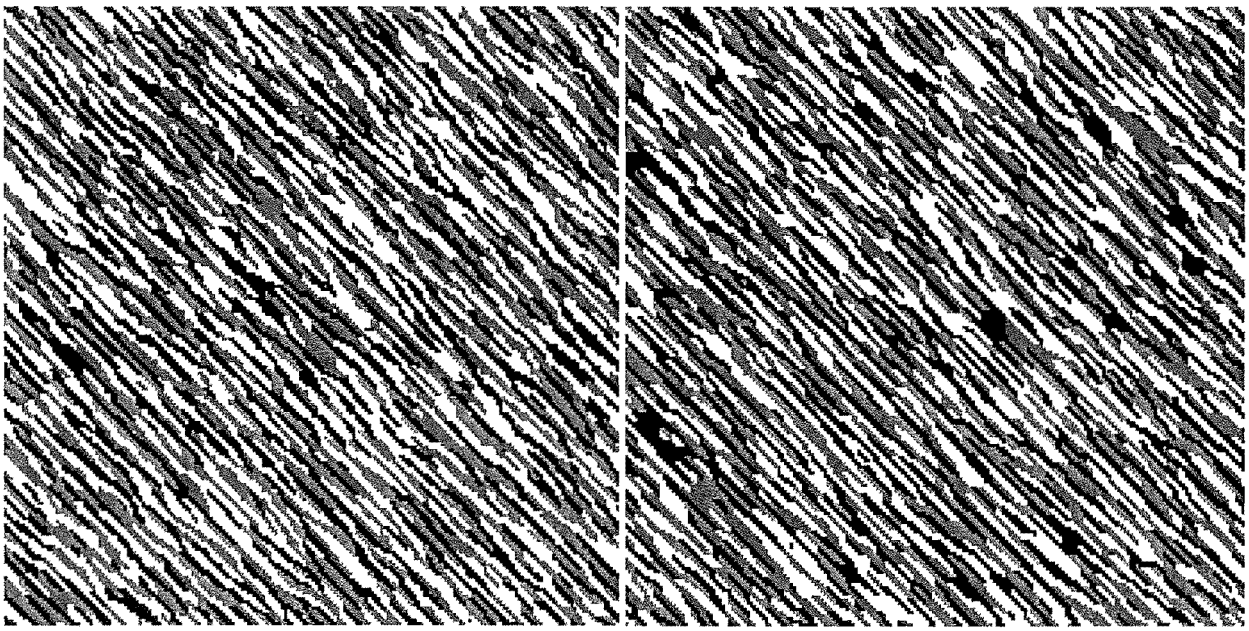
(a) $\Theta = [2\ \ 1.5\ \ 0.7\ \ 0.4]$ (b) $\Theta = [1\ \ 2\ \ 0.9\ \ 2]$

(c) $\Theta = [1\ \ 2\ \ 2\ \ 0.9]$ (d) $\Theta = [2\ \ 2\ \ 1\ \ 0.5]$

Figure 10.6 Realisations of a GRF. Images are of the size 200 × 200 pixels and contain 4 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the eight-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [2\ \ 1.5\ \ 0.7\ \ 0.4]$, (b) $\Theta = [1\ \ 2\ \ 0.9\ \ 2]$, (c) $\Theta = [1\ \ 2\ \ 2\ \ 0.9]$ and (d) $\Theta = [2\ \ 2\ \ 1\ \ 0.5]$. The parameter vector is given by $\Theta = [\ |\ \ -\ \ \backslash\ \ /\ ]$ where the lines indicate the clique type (| for example is the vertical clique).



(a) $\Theta = [1\ \ 2\ \ 3\ \ 4]$



(b) $\Theta = [1\ \ 1\ \ 1\ \ -1]$ (c) $\Theta = [2\ \ 2\ \ 2\ \ -2]$

Figure 10.7 Realisations of a GRF. Images are of the size 200 × 200 pixels and contain 3 classes. Fifty iterations (complete sweeps) were used to generate these using the Metropolis sampler. The images are based on the eight-neighbourhood with the parameter vectors for the cliques as follows: (a) $\Theta = [1\ \ 2\ \ 3\ \ 4]$, (b) $\Theta = [1\ \ 1\ \ 1\ \ -1]$ and (c) $\Theta = [2\ \ 2\ \ 2\ \ -2]$. The parameter vector is given by $\Theta = [\,|\ \ -\ \ \backslash\ \ /\,]$ where the lines indicate the clique type (| for example is the vertical clique).

10.1.2 The parameter estimation on the GRF realisations

The new parameter estimation method as described in Chapter 8 was used to estimate the GRF parameters Θ of the clique potential functions. For the parameter estimation, the number of classes and the correct neighbourhood system were given. Different clique types were assigned the same parameter if they were expected to have the same parameter value. For example [{ | −}{\ /}] means that the vertical and the horizontal clique have the same parameter value assigned to them, and similarly for the two diagonal clique types. The value of $\varepsilon = 1/(n_{\text{equi}} L^{N+1})$, where n_{equi} is the number of equivalencies, was used in the cost function (Equation 8.14) throughout all experiments in the thesis. The cost function of the new parameter estimation method in Equation 8.14 was minimised using MATLAB's `fminsearch` function.

For comparison Besag's coding method was also used to estimate the parameters from the GRF realisations making the same assumptions for the model type as above. The optimisation was performed in the same way using MATLAB's `fminsearch`. As an initial estimate (starting point of the estimation) all parameter values were set to one.

Table 10.1 gives the parameter vectors estimated with the new parameter estimation method and for comparison the parameters estimated with Besag's coding method. The neighbourhood system, number of classes and the way the cliques were combined are also given in this table.

For most GRF realisations the convergence to stable parameter values during the optimisation was rapid. In some cases for the GRF models with the eight-neighbourhood system and four different parameter values, the convergence was less rapid but still settled on a stable value. This is due to the higher dimensionality of the optimisation problem. This happened for both parameter estimation methods.

Table 10.1 indicates that the new parameter estimation method performs slightly better than Besag's parameter estimation method. Both methods show similar behaviour in that most parameter values are close to the expected parameter values that have been used to create the GRF realisations. In most cases the parameters are underestimated. As far as the new method of parameter estimation is concerned the reason for this may be the small sample size. The GRF realisations are not large enough to give a good representation of a GRF model especially for a more complex GRF model. Larger neighbourhood sizes and different parameters for the individual cliques create higher numbers of equivalencies and therefore more samples are needed to represent the local joint density. Since the creation of the GRF realisations and the estimation of the GRF model parameters is based on stochastic processes some deviation is to be expected. Finally it has to be pointed out that there is no independent way of evaluating the quality of the GRF realisations.

Neigh-bourhood system	Number of classes	Parameter vector used to create the GRF realisation	Parameter model	Parameters estimated with the new parameter estimation method	Besag's estimated parameters
4	3	[0.3 0.3]	[{ | −}]	[0.30]	[0.30]
	3	[0.6 0.6]		[0.60]	[0.55]
	3	[1.0 1.0]		[0.86]	[0.90]
	3	[2.0 2.0]		[0.99]	[1.32]
	3	[0.3 0.15]	[| −]	[0.30 0.17]	[0.30 0.15]
	3	[0.6 0.3]		[0.59 0.30]	[0.58 0.29]
	3	[0.3 −0.15]		[0.30 −0.14]	[0.31 −0.14]
	3	[0.6 −0.3]		[0.58 −0.28]	[0.58 −0.30]
	4	[0.3 −0.3]		[0.28 −0.32]	[0.30 −0.32]
	4	[0.6 −0.6]		[0.60 −0.59]	[0.60 −0.59]
	4	[1.0 −1.0]		[1.05 −0.87]	[1.01 −1.00]
	4	[2.0 −2.0]		[1.82 −1.06]	[1.68 −2.01]
8	3	[2 2 1 1]	[{ | −}{\ /}]	[1.16 0.57]	[1.11 0.66]
	3	[1 1 0.5 0.5]		[0.78 0.61]	[0.81 0.45]
	3	[0.5 0.5 0.25 0.25]		[0.52 0.25]	[0.52 0.25]
	3	[0.25 0.25 0.12 0.12]		[0.25 0.13]	[0.25 0.13]
	3	[2 2 −1 −1]		[1.98 −1.00]	[1.84 −0.94]
	3	[4 4 −2 −2]		[3.13 −1.61]	[3.22 −1.69]
	3	[−1 −1 2 2]		[−0.77 1.37]	[−0.77 1.37]
	3	[−2 −2 4 4]		[−1.02 1.81]	[−0.85 1.40]
	4	[2 1.5 0.7 0.4]	[| − \ /]	[1.58 1.20 0.64 0.31]	[1.28 1.00 0.40 0.24]
	4	[1 2 0.9 2]		[0.69 1.62 0.50 1.06]	[0.62 1.03 0.43 0.85]
	4	[1 2 2 0.9]		[0.60 1.21 1.03 0.58]	[0.53 1.04 0.91 0.51]
	4	[2 2 1 0.5]		[1.26 1.31 0.59 0.28]	[1.17 1.31 0.61 0.29]
	3	[1 2 3 4]		[0.42 0.74 1.48 1.52]	[0.38 0.65 1.18 1.29]
	3	[1 1 1 −1]		[0.96 0.94 0.96 −0.94]	[0.98 0.98 0.99 −0.95]
	3	[2 2 2 −2]		[1.68 1.74 1.59 −1.62]	[2.06 2.10 1.63 −2.19]

Table 10.1 This table gives the GRF model type and the model parameters used to create the GRF realisations. Figure 10.1 through to Figure 10.7 show the corresponding GRF realisations used in the parameter estimation. Further the estimated parameter vectors using the new parameter estimation method and Besag's coding method are given. The parameter model indicates the clique types that had the same parameter value assigned to it during the parameter estimation.

10.2 PARAMETER ESTIMATION FROM HAND DRAWN AND MEASURED DATA

In the previous section GRF realisations were used to validate the parameter estimation method. In this section the parameter estimation is applied to data that is not a GRF realisation. The difficulty with this is that the true underlying optimal parameter values are not known. Therefore in order to assess the quality of the estimated parameters, a MAP classification on the same data with noise is performed using the estimated parameters in the prior GRF model. As in the previous section, Besag's parameter estimation is used to compare results.

10.2.1 The test data

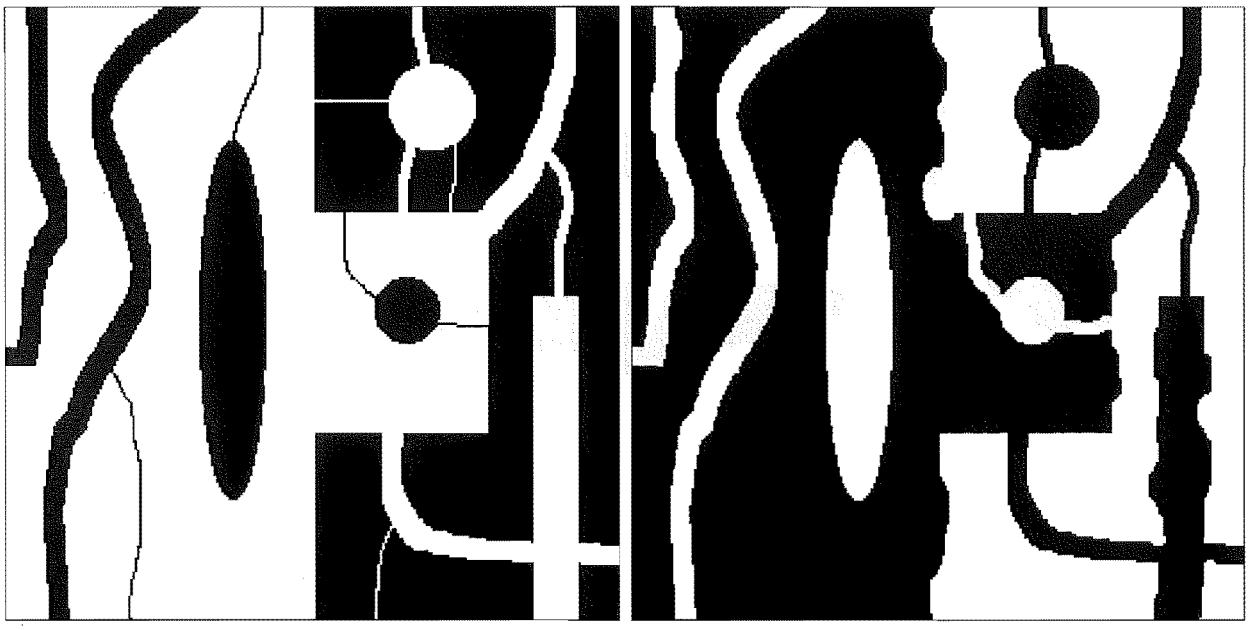
The test data is in the form of label images. The first test label images in Figure 10.8(a) and Figure 10.8(b) are hand drawn images with two classes indicated by the black and white areas. Both images have the size of 256×256 pixels. Figure 10.8(b) is a modified version of Figure 10.8(a): labels values have been exchanged, small line objects have been erased and there are less straight boundaries between the two classes. The first scene shown in label image Figure 10.8(a) is referred to as `LHALF1` and the second scene shown in the label image Figure 10.8(b) as `LHALF2`.

The third test label image in Figure 10.9 is data generated from measured Geographic Information Systems (GIS) data indicating different types of ground cover of a scene of forestry blocks in Kaingaroa, New Zealand (see Figure 10.11). Five different classes of ground cover are present in the label image. Ordered from black to white they are: bare ground (black), Douglas Fir trees, Radiata Pine trees younger than 4 years, Radiata Pine trees from 4 years to 10 years in age and Radiata Pine trees 10 years and older (white). The label image has the size of 512×512 pixels and is referred to as `KAINGAROA`.

10.2.2 The parameter estimation

The GRF model parameters were estimated for all three test label images shown in Figures 10.8(a), 10.8(b) and 10.9 using the newly developed parameter estimation method described in Chapter 8. The four- and the eight-neighbourhoods with different clique configurations were used as GRF models. Besag's parameter estimation method was used for comparison.

The results of the parameter estimation are in Table 10.2 and show some degree of similarity between the new parameter estimation method and Besag's parameter estimation method. The new parameter estimation method estimates smaller parameters, in most cases, for the `LHALF1` and `LHALF2` test label images. The most noticeable exception for `LHALF1` is the last model configuration (eight-neighbourhood with [| $-$ \ /] as clique configuration). Here a value of 2.60 has been estimated using the new parameter estimation method for the vertical clique type. From experience with GRF models the value of 2.60 in the parameter vector is too large to be a realistic estimate for this type of label image. The parameter estimates from Besag's parameter estimation method for `LHALF2` are all larger than the estimated parameter values from the new parameter estimation



(a) LHALF1 (b) LHALF2

Figure 10.8 Two hand drawn label images of size 256 × 256 pixels. Both figures have two classes indicated by the black and the white areas. Figure (a) is referred to as LHALF1 and Figure (b) as LHALF2.

method. Especially the second and fifth model configuration of LHALF2 (four-neighbourhood with [| −] and the eight-neighbourhood with [| − \ /] as clique configuration) where the vertical clique has a parameter valuer greater than 8. These values are also considered too large to be realistic estimates.

For the KAINGAROA label image the estimated parameters of the two different parameter estimation methods are very similar. The new parameter estimation method estimates larger values in this case especially for the eight-neighbourhood configurations. However in this example no obviously unrealistic parameter values are present.

10.2.3 Quality of the parameter estimation

In order to assess the quality of the estimated parameter vectors, the parameters were used with their corresponding GRF model to generate MAP classifications on noisy data corresponding to the label test images. The difference between the label test images and the MAP classification results give an indication of the quality of the estimated parameters, assuming that better prior model parameters will result in better classifications.

For both hand drawn label images, LHALF1 in Figure 10.8(a) and LHALF2 in Figure 10.8(b), single-look intensity speckle noise was simulated. Therefore each class was assigned corresponding data from a negative exponential distribution for which the mean value was the only parameter of the distribution (see Section 4.1 for details on this speckle model). In this case the white areas received a mean intensity value of 8 and the black areas received a mean intensity value of 4.

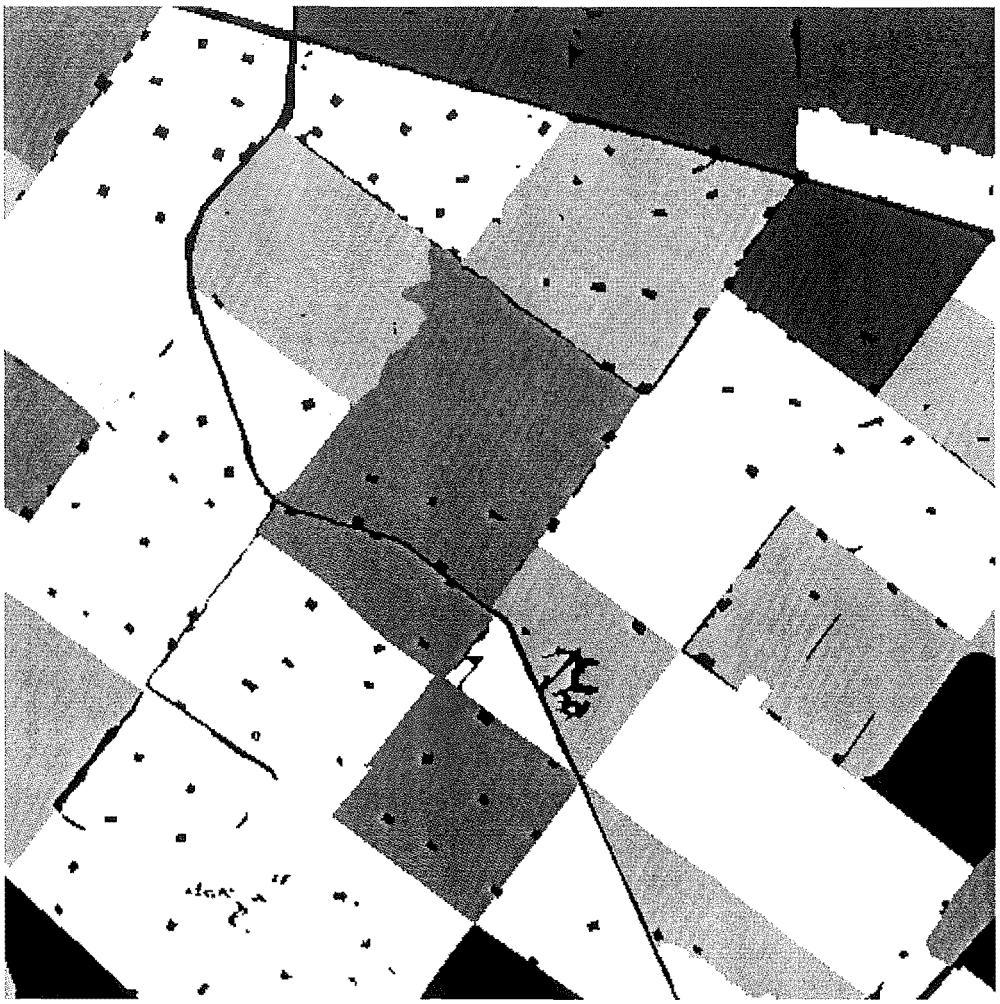
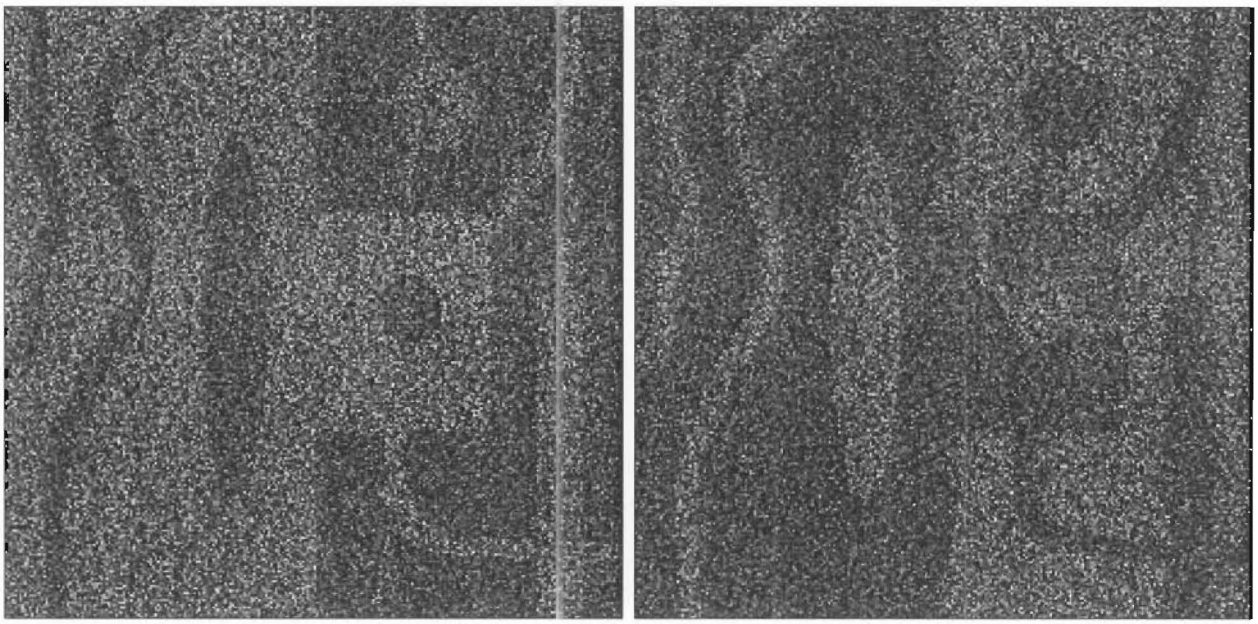


Figure 10.9 A label image created from GIS data indicating different types of ground cover of a scene of forestry blocks in Kaingaroa New Zealand (see Figure 10.11). Five different classes of ground cover are present in the image, ordered from black to white they are: bare ground (black), Douglas Fir trees, Radiata Pine trees younger than 4 years, Radiata Pine trees 4 years and older but younger than 10 years, Radiata Pine trees 10 years and older (white). The image has the size of 512×512 pixels and is referred to as `KAINGAROA`.

Neighbour-hood system	Parameter model	Parameters estimated with the new parameter estimation method	Besag's estimated parameters
LHALF1, with two classes (see Figure 10.8(a))			
4	[{ | −}]	0.88	1.20
4	[| −]	[1.16 1.05]	[1.36 0.94]
8	[{ | − \ /}]	0.49	0.49
8	[{ | −}{\ /}]	[1.24 0.05]	[1.24 −0.10]
8	[| − \ /]	[2.60 1.00 0.20 0.29]	[1.32 1.06 −0.10 −0.05]
LHALF2, with two classes (see Figure 10.8(b))			
4	[{ | −}]	0.90	1.91
4	[| −]	[1.02 1.18]	[8.78 1.48]
8	[{ | − \ /}]	0.51	1.14
8	[{ | −}{\ /}]	[0.95 0.32]	[1.68 0.59]
8	[| − \ /]	[1.54 1.04 0.43 0.40]	[8.50 1.25 0.68 0.55]
KAINGAROA ground cover, with five classes (see Figure 10.9)			
4	[{ | −}]	1.12	1.13
4	[| −]	[1.29 1.22]	[1.23 1.04]
8	[{ | − \ /}]	0.82	0.63
8	[{ | −}{\ /}]	[0.96 0.85]	[0.64 0.62]
8	[| − \ /]	[0.91 0.92 0.73 0.68]	[0.68 0.61 0.57 0.67]

Table 10.2 Results from the parameter estimation of the three label test images shown in Figures 10.8(a), 10.8(b) and 10.9 for different neighbourhoods and clique configurations as indicated in columns one and two. Columns three and four give the estimated parameter vectors Θ for the new parameter estimation method and for Besag's parameter estimation method respectively.



(a) LHALF1 (b) LHALF2

Figure 10.10 These figures show simulated single-look intensity SAR data for the hand drawn binary label images shown in Figure 10.8(a) and in Figure 10.8(b). The first class indicated by the black areas in the label images is modelled as a negative exponential with a mean intensity value of 4 and the second class indicated by the white areas as a negative exponential with a mean intensity value of 8. Note that the images have been gamma scaled for better display.

Figures 10.10(a) and 10.10(b) show the simulated single-look intensity images of the hand drawn label test images.

For the scene KAINGAROA in the label image Figure 10.9, measured SAR data is available. The SAR data is in the form of multiband (C, L and P) polarimetric (HH, HV and VV) 9-look intensity AIRSAR data[1]. Each pixel corresponds to a 4.63m × 3.33m square on the ground. Figure 10.11 shows the HH intensity data of the C-band. Note that the image is gamma scaled in order to reduce the dynamic range of the data for better display.

For the classification the MAP estimate was obtained using simulated annealing as described in Chapter 6. The MAP density was made up of the likelihood function and the prior model. In this case the likelihood function modelled the label value with regard to the noisy data and the prior model rendered some label images more likely than others. The GRF model was used as the prior model. Since the aim in this classification is to assess the quality of the GRF model parameters, the likelihood functions and their parameters are chosen to be as close to optimal as possible.

The negative exponential density, using the correct means as given, was used as the likelihood function for the two single look SAR data test images (from LHALF1 and LHALF2). This means that the same density and parameters were used in the model as were used to create the noisy data.

[1]Data obtained from Landcare Research, New Zealand. The AIRSAR system is a side-looking imaging radar system aboard a modified NASA DC 8 aircraft and first flew in 1988 as part of NASA's Earth Science Enterprise. AIRSAR was designed and built by the Jet Propulsion Laboratory (JPL).

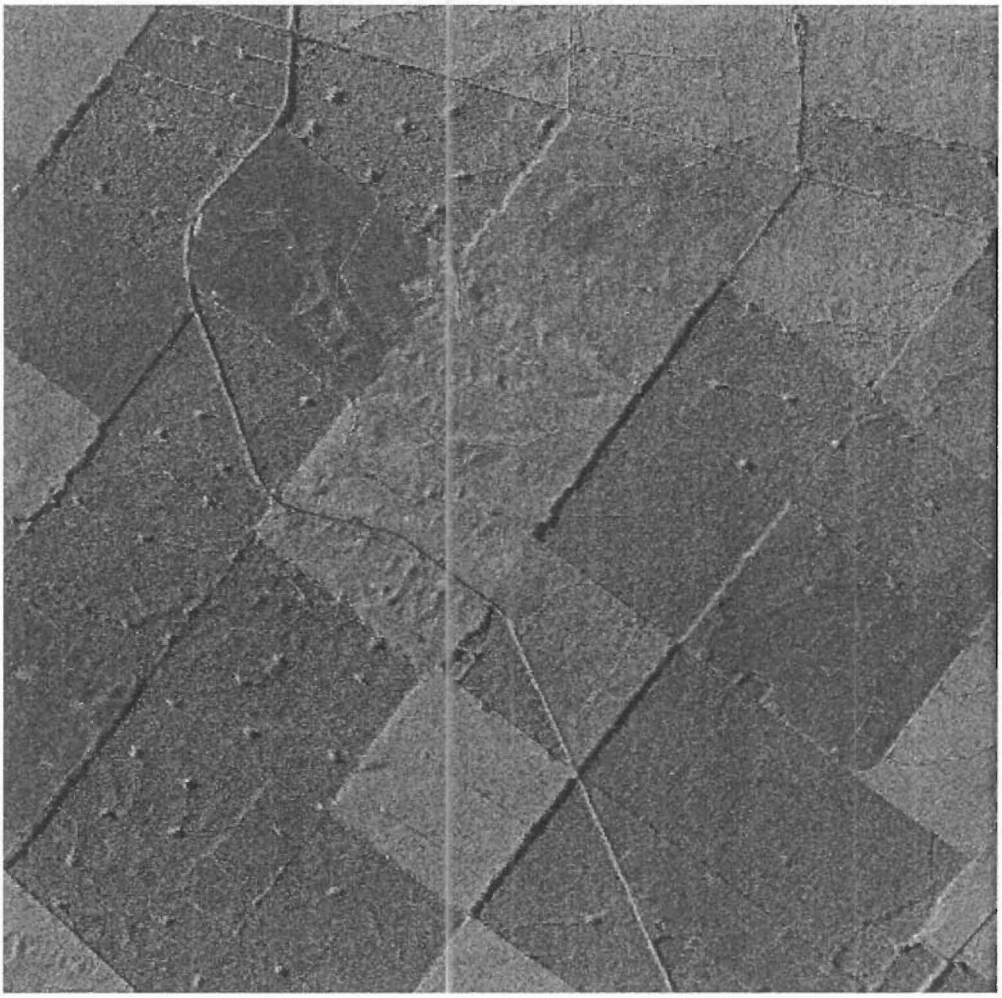


Figure 10.11 This image shows the gamma scaled HH intensity data of the C-band from 9-look, 3-band (C, L and P) polarimetric AIRSAR data. The scene shows forestry blocks in the Kaingaroa area New Zealand in August 2000. Figure 10.9 shows the underlying GIS data indicating the ground cover. The image has the size of 512×512 pixels with each pixel corresponding to a 4.63m $\times$ 3.33m square on the ground.

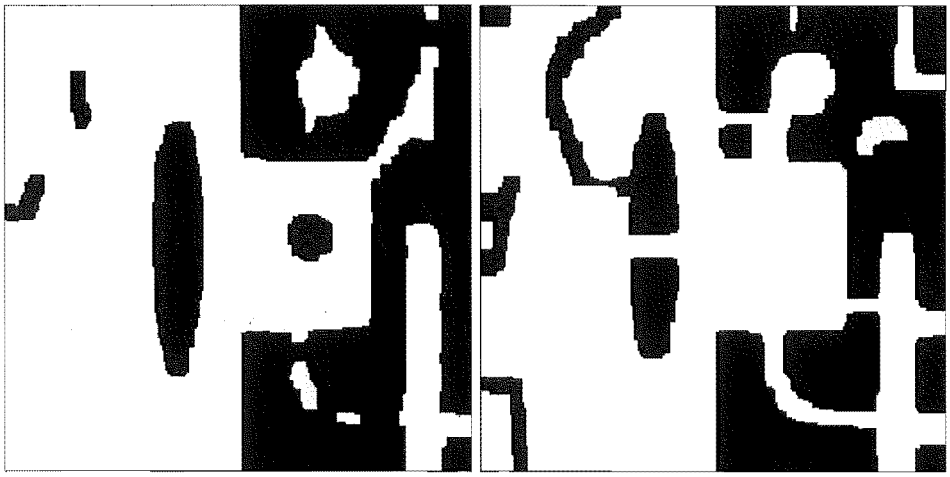
A nine dimensional Gaussian density was used for the AIRSAR data of the KAINGAROA scene, with one dimension for each band and polarisation intensity (intensity of the C-band HH polarisation for example). Single channel multilook intensity SAR data is best described by a chi-square density as shown in Section 4.1. The nine dimensional Gaussian density however is a good model in this case since the number of looks (nine) is high enough to give the chi-square density a Gaussian shape and correlations between all nine channels are taken into account. For each class (bare ground (black), Douglas Fir trees, Radiata Pine trees younger than 4 years, Radiata Pine trees 4 years and older but younger than 10 years, Radiata Pine trees 10 years and older (white) as indicated in Figure 10.9) a mean vector of length nine and a 9×9 covariance matrix was estimated using the corresponding AIRSAR data. For example, all AIRSAR data for the Douglas Fir trees was extracted using the label image Figure 10.9 to mask out the rest of the AIRSAR data. Areas close to the edge of boundaries between classes were not used since they are likely to have been corrupted by boundary effects such as shadows. Then the mean vector and the covariance matrix were estimated from that data and assigned to the class representing the Douglas Fir trees.

The simulated annealing method to find the MAP estimate was performed according to the description in Section 6.1. The algorithm was programmed in C and run in a linux environment. 1000 iterations (complete sweeps) with the cooling constant $C_A = 3$ were performed. For each noisy data set (see Figures 10.10(a), 10.10(b) and 10.11) five different GRF model configurations were used. For each model configuration two parameter estimates were tested, one obtained using the new parameter estimation method and the other using Besag's parameter estimation method. For all of the model configurations and parameter sets used see Table 10.2.

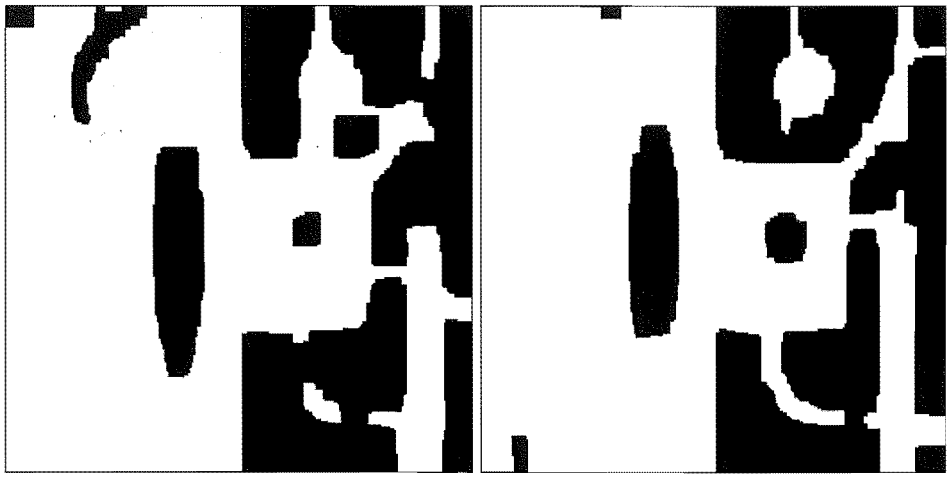
The classification results for all three test data sets with the parameter sets from both parameter estimation methods and the different model configurations are shown in Figures 10.12 - 10.18. The results using the parameters from the new parameter estimation method are shown on the left and the ones using Besag's parameter estimation method on the right. Table 10.3 gives the percentage of wrongly classified area for all cases. For comparison the maximum likelihood (ML) estimate has been calculated and the results are shown in Figures 10.20(a), 10.20(b) and 10.21. For the ML estimate no prior model was used and the MAP estimate reduces to the ML estimate (this is the same as setting the parameter vector Θ to zero).

Table 10.3 shows that for the LHALF1 classification results Besag's estimated parameters for the last two model configurations (eight-neighbourhood with [{ | −}{\ /}] and [| − \ /] as clique configuration) perform better than the parameters from the new parameter estimation method. These parameter sets have been overestimated by the new parameter estimation method resulting in many smaller features not being detected and boundaries being shifted by the classification. However, in the second example LHALF2 all classification results with the parameters from the new parameter estimation method outperform the results using Besag's estimated parameters. It is clear that the parameters have been overestimated with Besag's parameter estimation method.

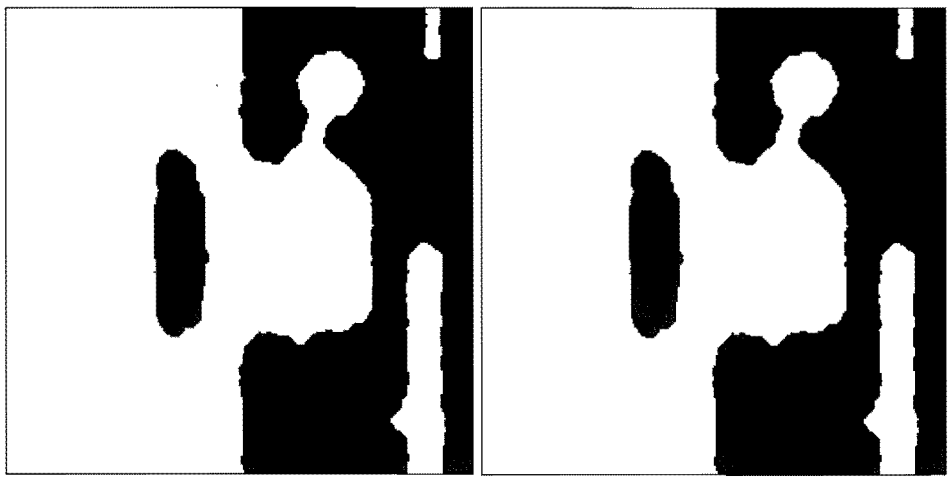
It is interesting to note that, for this test data (LHALF2), the results get worse as the complexity of the model increases. For optimal parameters the opposite would be expected (a higher complexity



(a) $\Theta = [0.88]$ ([{ | −}]) (b) $\Theta = [1.20]$ ([{ | −}])

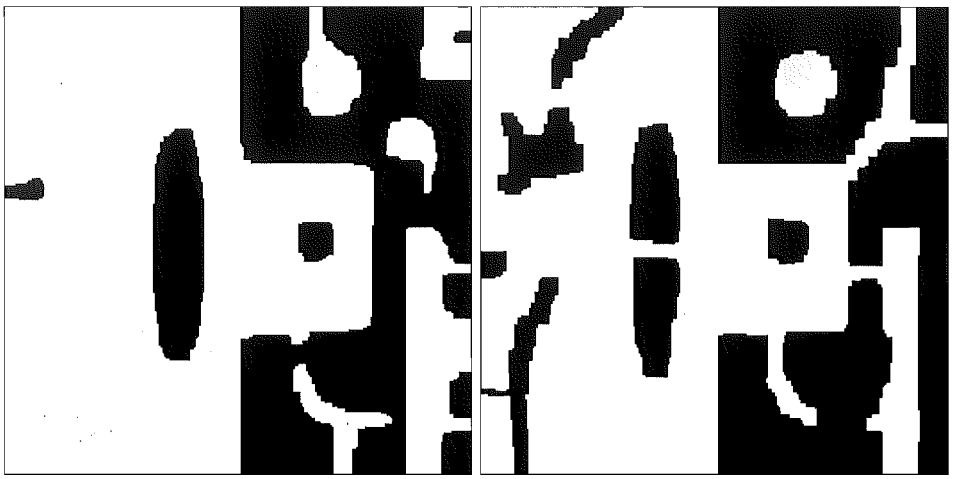


(c) $\Theta = [1.16\ \ 1.05]$ ([| −]) (d) $\Theta = [1.36\ \ 0.94]$ ([| −])



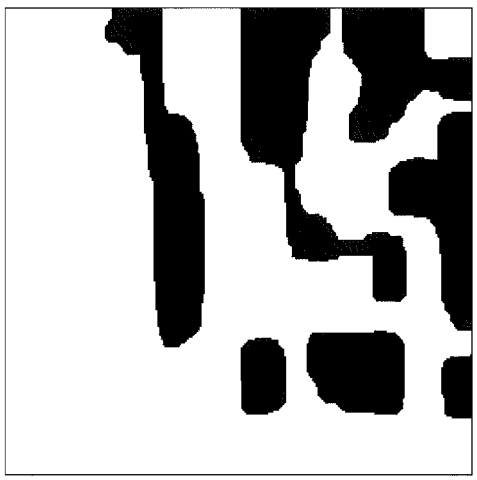
(e) $\Theta = [0.49]$ ([{ | − \ /}]) (f) $\Theta = [0.49]$ ([{ | − \ /}])

Figure 10.12 These figures show the classification result from the noisy `LHALF1` data (see Figure 10.10(a)). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.

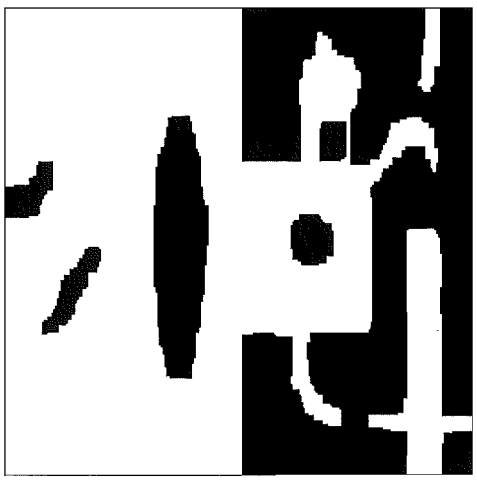


(a) $\Theta = [1.24 \;\; 0.05\,]$ ([{ | −}{\ /}])

(b) $\Theta = [1.24 \;\; -0.10\,]$ ([{ | −}{\ /}])

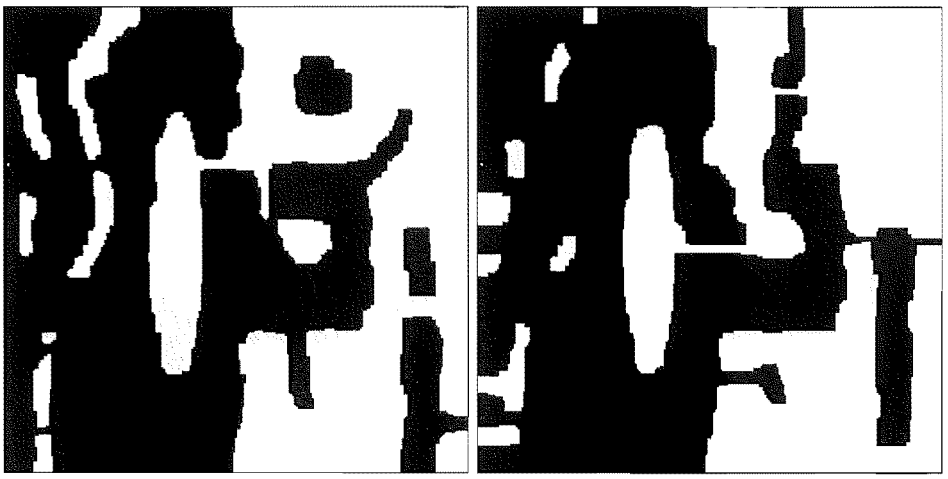


(c) $\Theta = [2.60 \;\; 1.00 \;\; 0.20 \;\; 0.29\,]$
([| − \ /])

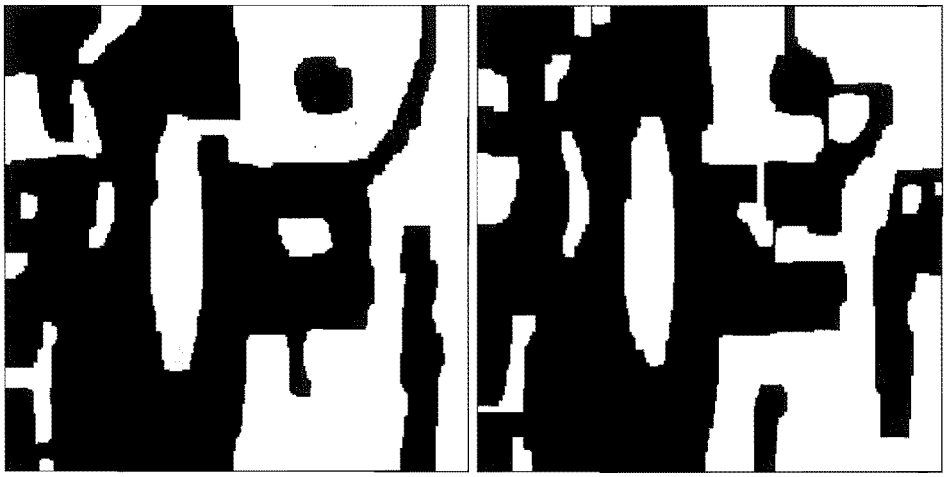


(d) $\Theta = [1.32 \;\; 1.06 \;\; -0.10 \;\; -0.05\,]$
([| − \ /])

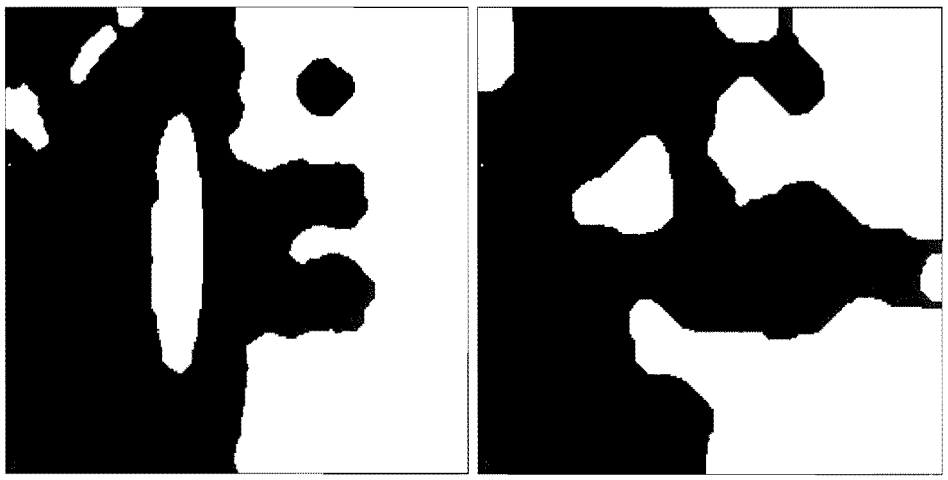
Figure 10.13 These figures show the classification result from the noisy `LHALF1` data (see Figure 10.10(a)). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.



(a) $\Theta = [0.90]$ ([{ | −}]) (b) $\Theta = [1.91]$ ([{ | −}])

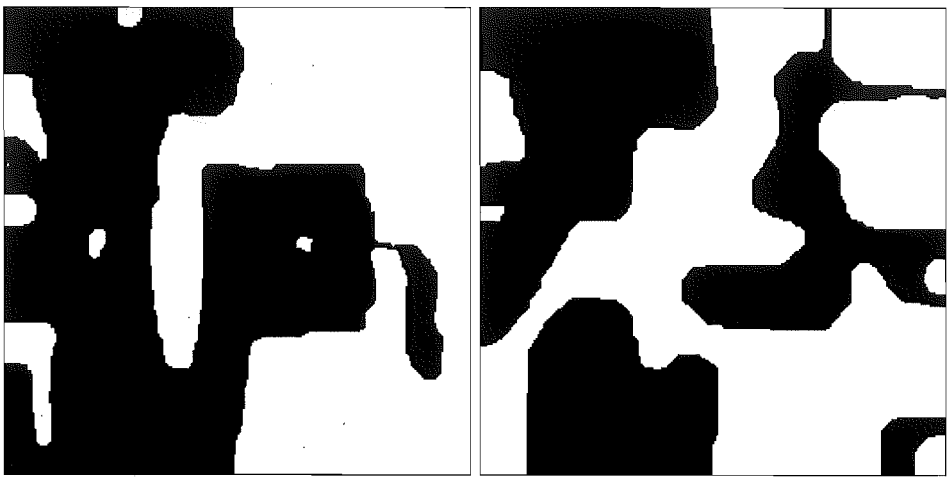


(c) $\Theta = [1.02\ \ 1.18]$ ([| −]) (d) $\Theta = [8.78\ \ 1.48]$ ([| −])



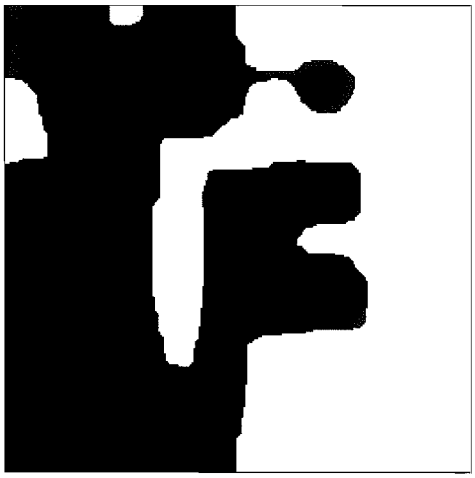
(e) $\Theta = [0.51]$ ([{ | − \ /}]) (f) $\Theta = [1.14]$ ([{ | − \ /}])

Figure 10.14 These figures show the classification result from the noisy LHALF2 data (see Figure 10.10(b)). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.

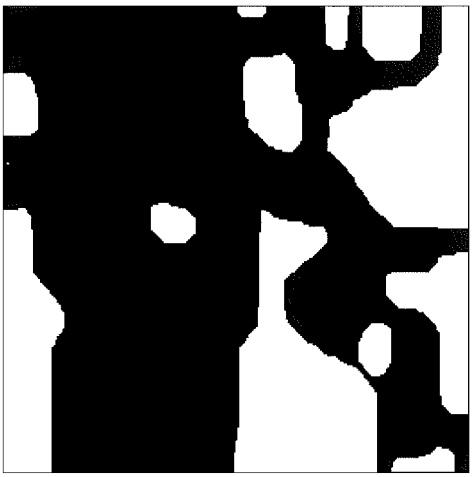


(a) $\Theta = [0.95\ \ 0.32\,]$ ([{ | −}{\ /}])

(b) $\Theta = [1.68\ \ 0.59\,]$ ([{ | −}{\ /}])

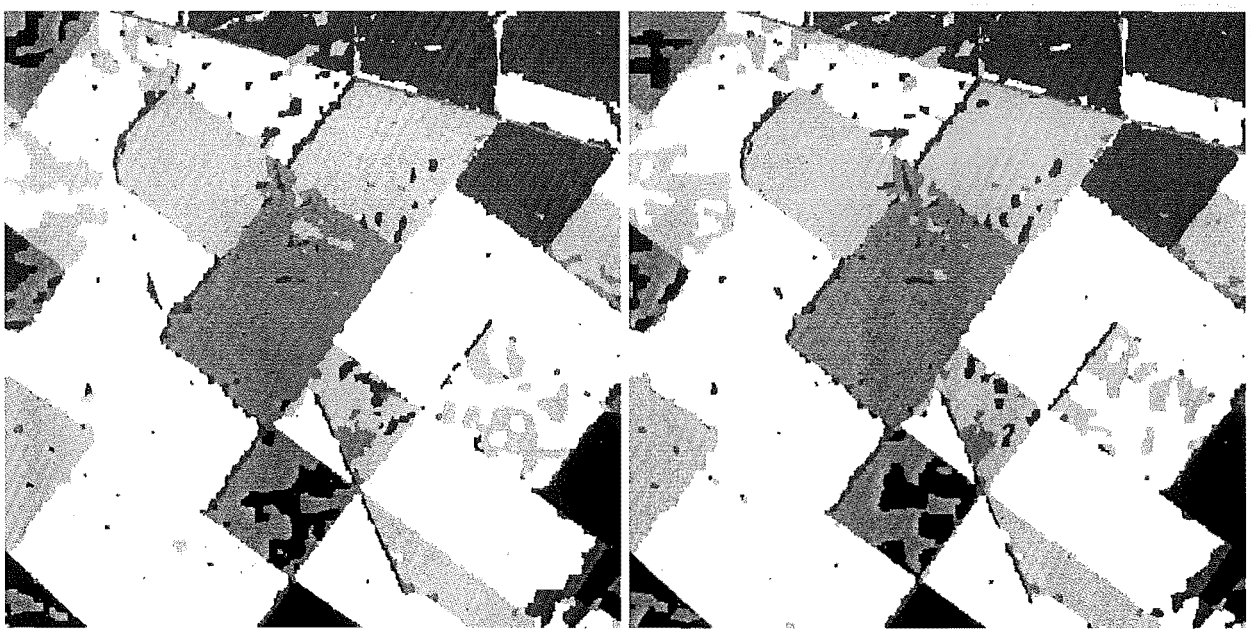


(c) $\Theta = [1.54\ \ 1.04\ \ 0.43\ \ 0.40\,]$
([| − \ /])



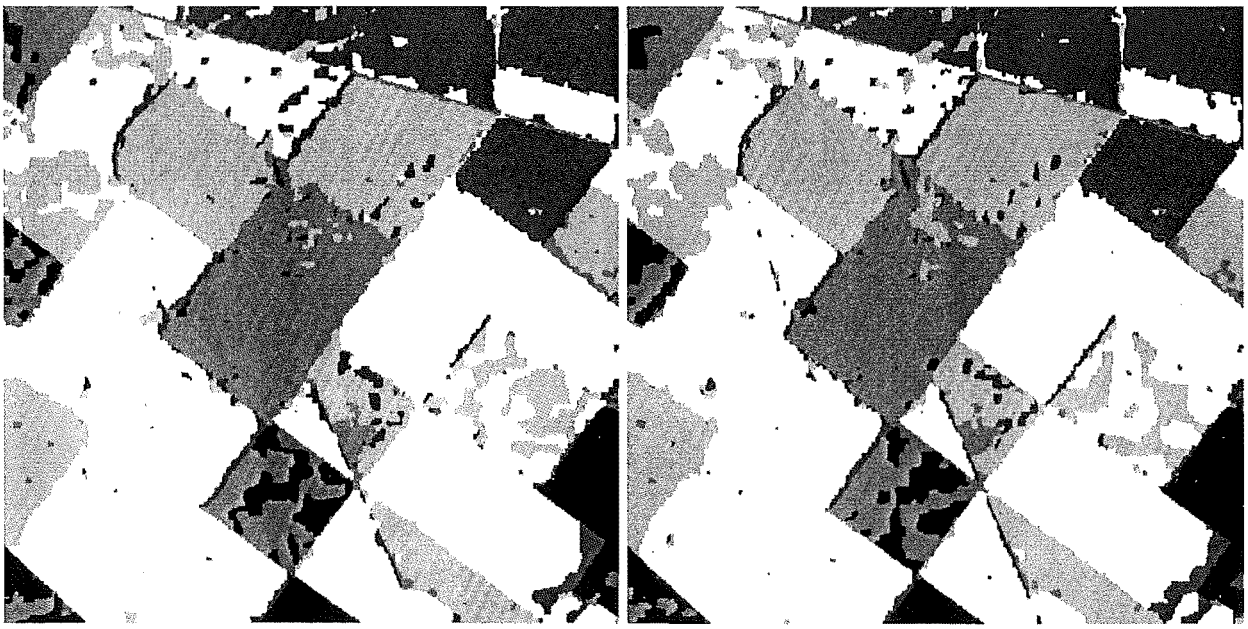
(d) $\Theta = [8.50\ \ 1.25\ \ 0.68\ \ 0.55\,]$
([| − \ /])

Figure 10.15 These figures show the classification result from the noisy `LHALF2` data (see Figure 10.10(b)). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.



(a) $\Theta = [1.12]$ ([{ | −}])

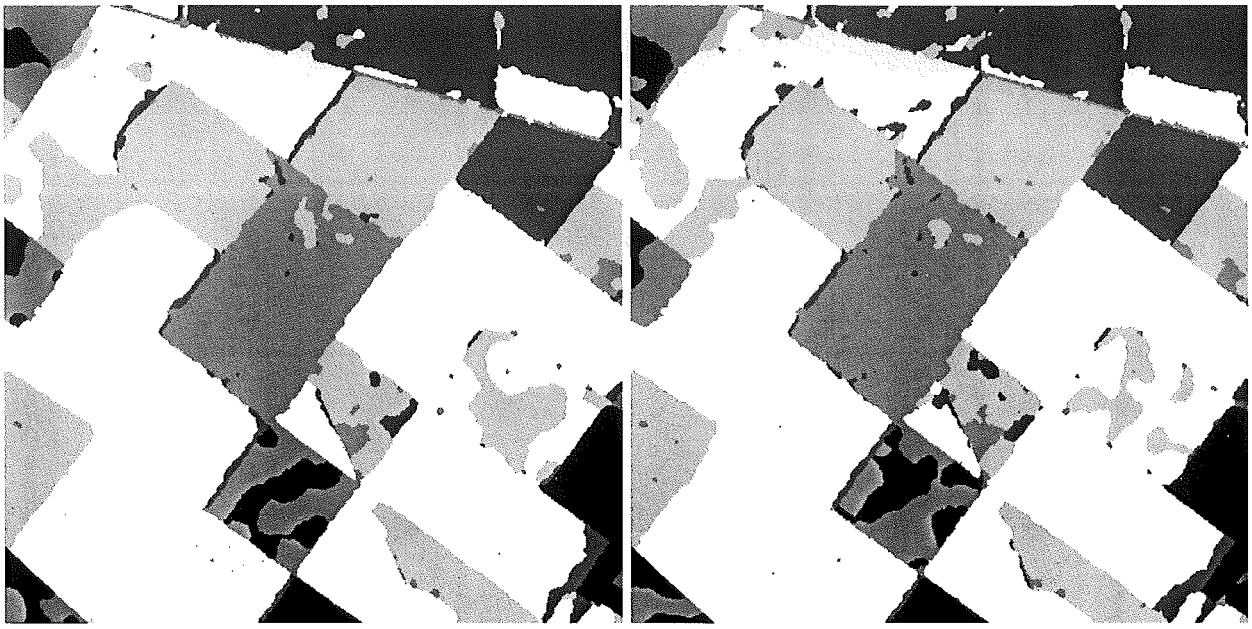
(b) $\Theta = [1.13]$ ([{ | −}])



(c) $\Theta = [1.29 \;\; 1.22]$ ([| −])

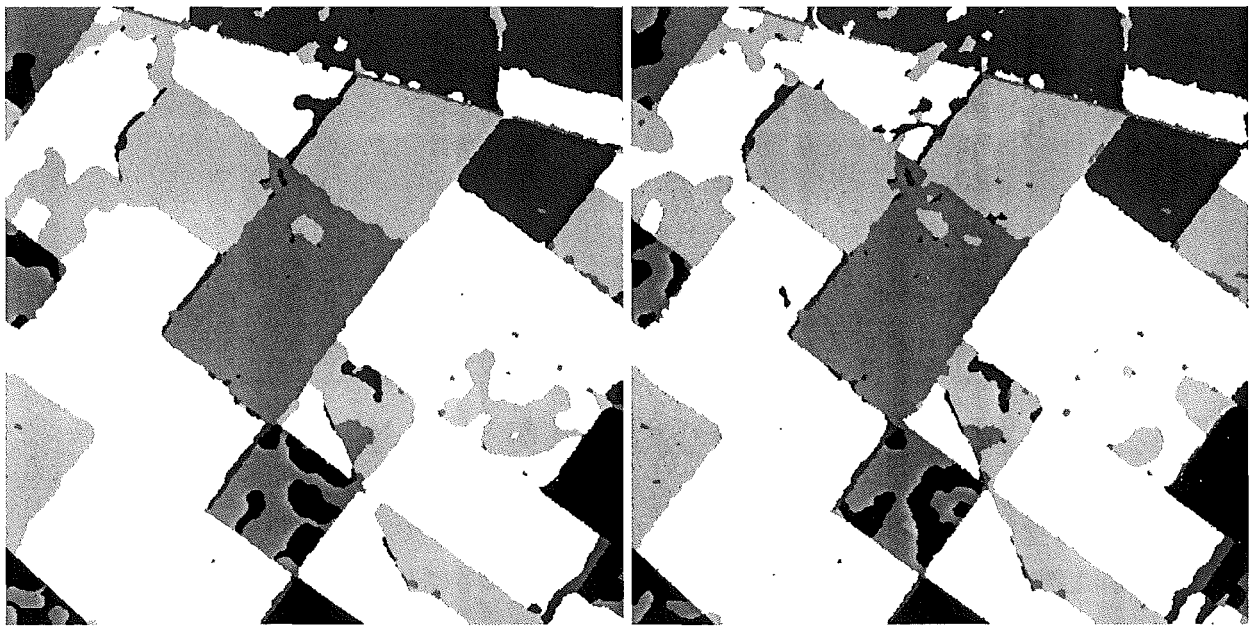
(d) $\Theta = [1.23 \;\; 1.04]$ ([| −])

Figure 10.16 These figures show the classification result from the KAINGAROA multi-band multi-polarisation intensity data (see Figure 10.11). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.



(a) $\Theta = [0.82]$ ([{ | -- \ /}])

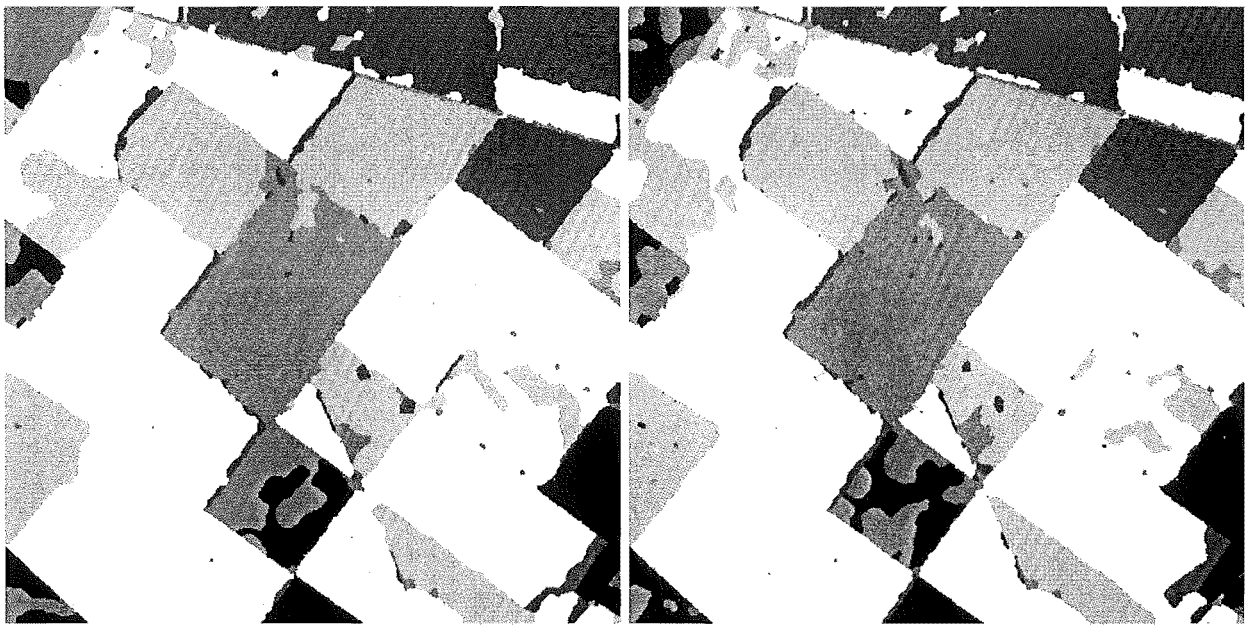
(b) $\Theta = [0.63]$ ([{ | - \ /}])



(c) $\Theta = [0.96\ \ 0.85]$ ([{ | -}{\ /}])

(d) $\Theta = [0.64\ \ 0.62]$ ([{ | -}{\ /}])

Figure 10.17 These figures show the classification result from the KAINGAROA multi-band multi-polarisation intensity data (see Figure 10.11). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.



(a) $\Theta = [0.91\ \ 0.92\ \ 0.73\ \ 0.68]$ ([| − \ /]) (b) $\Theta = [0.68\ \ 0.61\ \ 0.57\ \ 0.67]$ ([| − \ /])

Figure 10.18 These figures show the classification result from the KAINGAROA multi-band multi-polarisation intensity data (see Figure 10.11). For the figures on the left GRF model parameters, obtained with the new parameter estimation method, and on the right with Besag's parameter estimation method have been used. The parameters and the clique model are indicated below each figure.

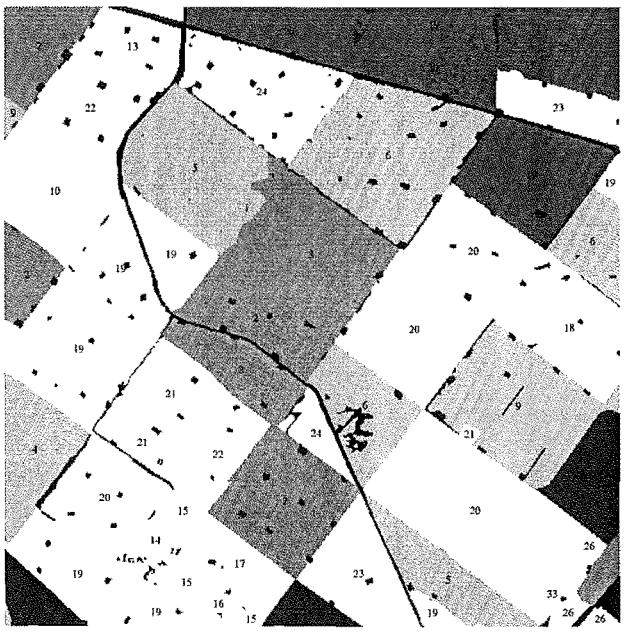
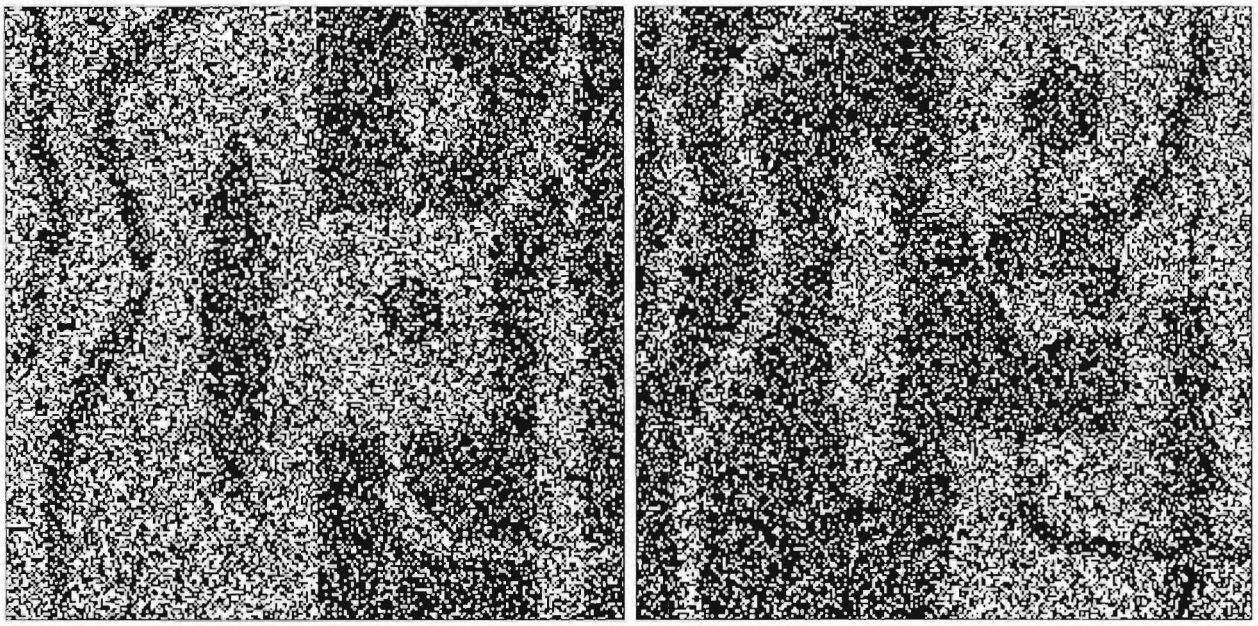


Figure 10.19 A label image created from GIS data indicating different types of ground cover of a scene of forestry blocks in Kaingaroa New Zealand (see Figure 10.11). Five different classes of ground cover are present in the image, ordered from black to white they are: bare ground (black), Douglas Fir trees, Radiata Pine trees younger than 4 years, Radiata Pine trees 4 years and older but younger than 10 years, Radiata Pine trees 10 years and older (white). The small numbers indicate the age of the trees in their surrounding area. The image has the size of 512×512 pixels.

Neighbourhood system	Parameter model	Percent of wrongly classified area with parameters estimated with the new parameter estimation method	Percent of wrongly classified area with Besag's estimated parameters
LHALF1, with two classes (see Figure 10.8(a))			
4	[{| −}]	9.17	10.60
4	[| −]	11.35	9.87
8	[{| − \ /}]	14.11	14.11
8	[{| −}{\ /}]	12.89	8.05
8	[| − \ /]	20.31	8.49
ML estimate (no prior model)		39.16	
LHALF2, with two classes (see Figure 10.8(b))			
4	[{| −}]	8.00	13.78
4	[| −]	9.02	13.52
8	[{| − \ /}]	16.45	23.90
8	[{| −}{\ /}]	17.13	24.19
8	[| − \ /]	19.09	27.74
ML estimate (no prior model)		35.91	
KAINGAROA ground cover, with five classes (see Figure 10.9)			
4	[{| −}]	16.59	16.68
4	[| −]	17.30	17.07
8	[{| − \ /}]	14.54	15.62
8	[{| −}{\ /}]	14.53	16.05
8	[| − \ /]	15.30	16.88
ML estimate (no prior model)		51.08	

Table 10.3 This table gives the percentage of wrongly classified area for the three test scenes (LHALF1, LHALF2 and KAINGAROA) with regard to the model type (neighbourhood system and clique configuration) and the parameter estimation method. For comparison the ML classifications (no prior model) have been added.



(a) LHALF1 (b) LHALF2

Figure 10.20 These figures show the ML classification result from the noisy LHALF1 and LHALF2 data (see Figure 10.10(a) and Figure 10.10(b)). For both figures no prior model was used in the classification, they therefore represent the ML estimate.

model should perform better than a simple model). This means that the parameter estimation results get less accurate as the complexity of the model increases. The reason for this is probably that a larger sample set (larger label image) is required to reliably estimate parameters for a complex model configuration than for a simple model configuration.

For the third example KAINGAROA the classification results for the two different parameter estimation methods are very similar. In most cases however the new parameter estimation method results in a better classification. The main source of misclassification is when the tree ages are very close to the cut off age for its class. For example two year old trees get mistaken partly for bare ground which is not surprising considering their size and spacing in a clear felled area. Another example is nine year old trees being classified as belonging to the above ten year old tree class. Figure 10.19 shows the label image KAINGAROA with numbers indicating the age in years of the trees in the forestry blocks. Therefore a large portion of the misclassification originates from the likelihood function and is not a result of the GRF model parameters of the prior model.

For comparison the classification from the ML estimates are shown in Figures 10.20(a), 10.20(b) and 10.21 and the percentages of wrongly classified area is given in Table 10.3. In this case no prior model is given and only the likelihood function is used to classify the noisy data. It is clear from inspection of the results that the MAP estimates well outperform the ML estimates in all cases.

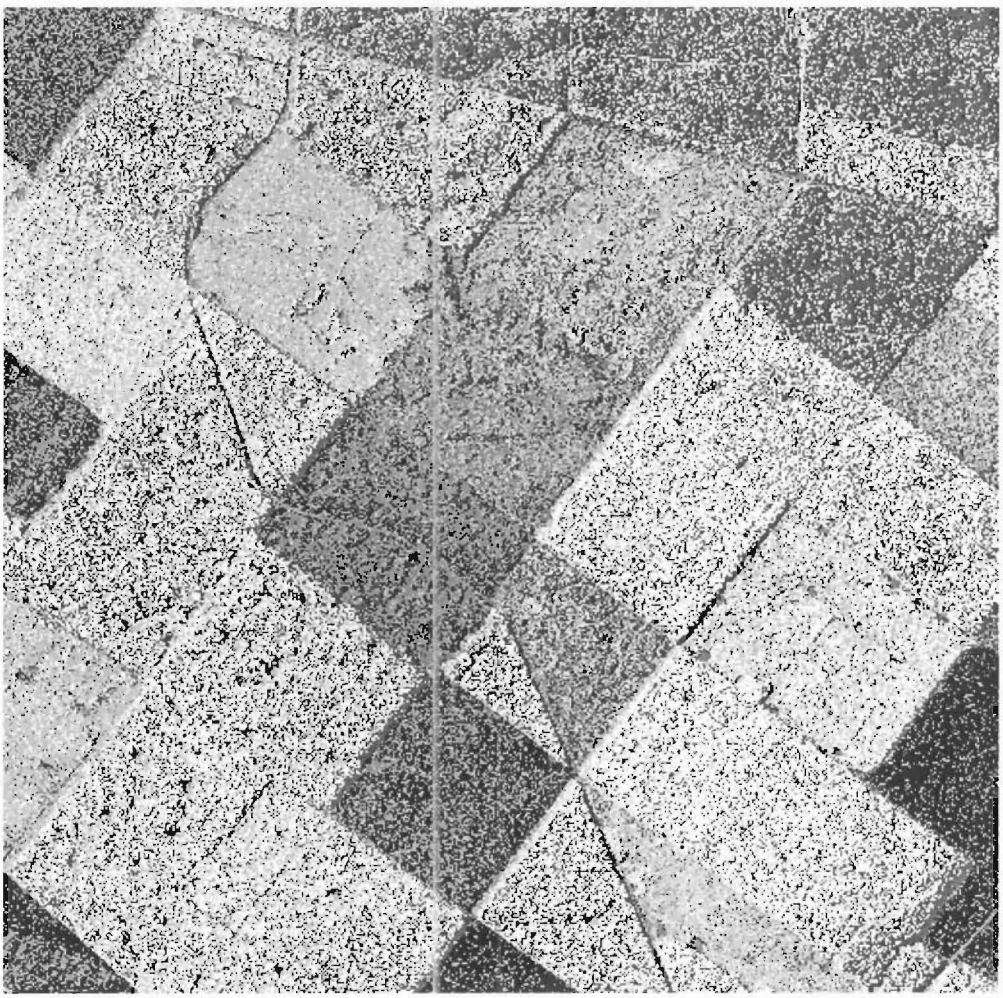


Figure 10.21 These figure shows the ML classification result from the KAINGAROA multi-band multi-polarisation intensity data (see Figure 10.11). No prior model was used in the classification, it therefore represent the ML estimate.

10.3 CONCLUSIONS

In this chapter the new parameter estimation method, introduced in Chapter 8, has been analysed and compared to Besag's parameter estimation method (coding method in Besag [1974]). Parameter estimation for GRF model parameters is an important part in developing an accurate prior model in a MAP segmentation or classification algorithm. In order to assess the quality of the new parameter estimation method two sets of experiments were performed.

In the first set of experiments GRF realisations were created using the metropolis sampler (see Section 5.10 for details). For the GRF realisations the model type, as in neighbourhood system and clique configuration, and the model parameters were known. The true model parameters can therefore be compared to the estimated ones. For 27 different GRF realisations, of size 200×200 pixels, the parameters were estimated using the new parameter estimation method and for comparison Besag's parameter estimation method. Comparing the results from the two parameters estimation methods to the true parameters, used in creating the GRF realisations, shows that both methods produce similar results (see Table 10.1). In many cases the new parameter estimation method produced better results than Besag's parameter estimation method. For both methods it is true that the higher the complexity of the GRF model (higher order neighbourhood and distinguishing between more clique types) and the larger the GRF model parameter values the less accurate are the parameter estimates. The reason for this is very likely the size of the GRF realisations. Larger sample sizes are required to represent more complex and more homogenous GRF model behaviour.

In the second set of experiments three test label images, two hand drawn with two classes and one extracted from a GIS data set with five classes, were used. The two hand drawn label images had the size of 256×256 pixels and the label image created from the GIS data had the size of 512×512 pixels. For all three test label images the GRF model parameters were estimated using different model types and clique configurations. For comparison Besag's method was used as well.

By just looking at the outcome of the parameter estimation from the three test label images of the two parameter estimation methods (see Table 10.2) it is clear that in some cases the parameters are estimated obviously too large (this can be concluded from experience with GRF models and their parameters). This was the case once for the new parameter estimation method (first test label image `LHALF1`, last model configuration, a value of 2.60) and twice for Besag's parameter estimation method (second test label image `LHALF2` second and last model configuration, values of larger than 8). It is not clear what the reason for this failure is.

The problem with this test data is that the true GRF model parameters are not known and it is therefore difficult to assess the quality of the parameter estimates, so noisy versions of the test label images were used in a MAP classification. For the two hand drawn test label images single look intensity SAR data was simulated. For the five class label image from the GIS data set true 9-look intensity AIRSAR data in the form of three bands (C, L and P) and each with three polarisations (HH, HV and VV) was available. Simulated annealing was then used to obtain the MAP estimates of the label images using all the different GRF model configurations and estimated

parameter sets. The MAP estimates were then assessed by comparing them with the original test label images and giving the percentage of wrongly classified area (see Table 10.3 for these values). The assumption was that better GRF model parameters would yield a better classification result.

For the first hand drawn test label image the percentage of wrongly classified area (see Table 10.3) indicates that both methods perform similarly, with slightly better results from Besag's parameter estimation method. For the second hand drawn test label image all results are better for the new parameter estimation method. For the third test label image, created from GIS data, the results indicate a slightly better fit for all but one of the parameter sets estimated using the new parameter estimation method.

For a further comparison the ML estimates (no prior model used) of the label images using the noisy test data were calculated. As expected these ML estimates gave results much worse than the worst MAP estimates. This indicates that using the prior model is advantageous no matter which of the two parameter estimation methods and what model configuration is used for these specific data sets.

It can be concluded that the new parameter estimation method generally outperforms Besag's although under certain circumstances one of the two parameter estimation methods may fail while the other still performs reasonably. If the test label image is small, a small neighbourhood and a single parameter for all cliques will yield better results rather than choosing a complex model configuration. The computational cost of the new parameter estimation method is slightly higher than Besag's parameter estimation method but both can be performed in a very short time. The simplicity of Besag's parameter estimation method is a clear advantage. However the mapping of equivalences (as described in Chapter 8) gives a deeper understanding of the underlying GRF model and may be useful in other applications.

Note that for both parameter estimation methods, the newly presented one and Besag's, only one sampling strategy (code) was used to obtain the estimate. Averaging the parameter estimates from all possible sampling strategies would improve the results in both cases.

Part IV

Discussion and conclusions

Chapter 11

DISCUSSION AND CONCLUSIONS

In this final chapter the results presented in the thesis are discussed and conclusions given. The first 6 chapters comprising Part I of the thesis provided the necessary background material for the development of the new contributions in the thesis. First in Chapter 2 the background of SAR (synthetic aperture radar) imaging systems is given. Then in Chapter 3 the required probability theory was introduced and in Chapter 4 the subsequent stochastic models for SAR image data were reviewed. Chapter 5 developed the theory of MRF (Markov random field) and GRF (Gibbs random field) models including the topics of sampling from a GRF and parameter estimation from a GRF. Chapter 6 combined the previous two chapters under the MAP (maximum *a posteriori*) estimation. The stochastic SAR models were used as the likelihood function and the GRF as the prior model. Two different optimisation algorithms to obtain the MAP estimate were given.

The theory of the contributions of this thesis were presented in Part II (Chapters 7 and 8) and evaluated and applied in Part III (Chapters 9 and 10). The contributions can be split into two main parts, the new likelihood function for multilook polarimetric intensity SAR data (Chapters 7 and 9) and the parameter estimation method for GRF models (Chapters 8 and 10). Therefore the discussion and conclusions are split in the same fashion.

11.1 THE NEW LIKELIHOOD FUNCTION FOR POLARIMETRIC INTENSITY SAR DATA

A closed-form expression for the joint density of intensity data from a three-polarisation n-look SAR system was derived in Section 7.1. This PDF (probability density function) describes the averaged intensity values from n samples from a complex three-dimensional zero-mean circular Gaussian random process. The derivation and the theoretical work related to this PDF are given in Chapter 7 while the experimental results are given in Chapter 9.

The main problem in deriving the new likelihood function was that three integrals had to be solved. In solving these integrals two cases had to be distinguished. In the first case with $n > 1$, results given by Miller et al. [1958] were used to solve the integrals. In the second case for $n = 1$ the derivation of Miller et al. [1958] breaks down and an alternate way of solving the integral was presented. The resulting expressions in both cases are in the form of an infinite series of modified

Bessel functions with seven independent parameters. The convergence of this series was proven in Section 9.1 on page 83 and experiments have shown that the convergence is rapid and no more than 20 terms in the series are needed.

In order to theoretically validate the new likelihood function, three special cases of this density were investigated in Section 7.2. In each case the joint density simplified to known expressions:

- **The uncorrelated case:** When the three intensity values were uncorrelated to each other (the covariance matrix is a diagonal matrix), the joint density collapsed to three independent chi-square distributions of the order $2n$.

- **The two-dimensional case:** In the case where one intensity value is independent of the other two, while the remaining two are correlated to each other, the joint density simplified to the product of a chi-square density and a two-dimensional PDF equivalent to the density of the two-polarisation n-look SAR data derived by Lee et al. [1994b].

- **The real case:** When there were no correlations between the imaginary and real parts of the complex signals, this case is equivalent to the sum of $2n$ intensity samples from a three-dimensional *real* zero-mean Gaussian distribution. Miller et al. [1958] derived the PDF for this case which matches the one given here.

In Section 7.3 the characteristic function of the new density was derived. The characteristic function provided a second approach to forming an expression for the joint density under investigation. This approach was followed through until only one integral was left which was considered too difficult to solve. However this expression for the joint density offers an alternative way of numerically evaluating the new likelihood function.

The parameters of the new likelihood function were defined as functions of the covariance matrix elements of the circular Gaussian distribution. In Section 7.4 a way of estimating these parameters from intensity sample data was derived. The parameters were expressed in terms of the moments of the intensity data with the help of the characteristic function which provided a convenient way to derive expressions for the moments of the new likelihood function.

In Section 9.2 the new likelihood function was validated by comparing it to simulated intensity data. The RMSE (root mean square error) values were used to show that simulated intensity data fits the new likelihood function well. For this experiment the exact parameters were used. These eight parameters of the new likelihood function were directly calculated from the covariance matrices used in producing the simulated data. In the following section (Section 9.3) these parameters were compared to parameters estimated from a small set of intensity sample data using the method derived in Section 7.4. Most of the estimated parameters corresponded closely with the actual parameter values. However each of the estimated parameters depended in varying degrees on higher order moments of the intensity sample data which was reflected in the variance of the estimates. It was shown that the estimated parameter with the highest variance had smaller influence on the new likelihood function compared to the other parameters and therefore a less accurate estimate

for this parameter was acceptable. It was also noted that an increasing sample size does provide more accurate estimates for all parameters.

In Section 9.4 the new likelihood function was compared to true SAR data. The chi-square goodness of fit test at a level of significance of $\alpha_{\text{sig}} = 0.05$ was used to test the hypothesis that the SAR sample data was from a population belonging to the new likelihood function. For 5 out of 9 selected SAR sample sets this hypothesis was rejected. Several reasons for failing the chi-square test were identified. The first one is the validity of the chosen SAR data with respect to the model. The ground cover of the SAR sample data has to be homogenous in the sense of radar reflectivity and texture free. However the complex covariance matrix of the data and the ENL values indicated that this was never completely true. Another problem was that only relatively small numbers of sample data could be collected making it more difficult to test the fit of the model.

For comparison the same chi-square test was performed on a second model. This model is the same as the new likelihood function but assumes independency (i.e., no correlation) between all three polarisations. All of the hypotheses were rejected and the chi-square test statistic values were greater than the ones from the new likelihood function. This clearly indicates that the new likelihood function fits the sample data better than the second model (i.e., the new likelihood function assuming independency between the polarisations).

The new likelihood function describes the true and uncorrupted speckle noise process of polarimetric multilook intensity data. For true SAR data, models have to be incorporated adding the effects of scene texture and other processes that effect the statistical nature of the signal. The new likelihood function can be used to separate the effect of speckle noise from other statistical processes present in the data and therefore give a better understanding of the underlying effects.

11.2 THE NOVEL PARAMETER ESTIMATION METHOD FOR GRF MODELS

In Chapter 8 a new method of estimating the parameters of a discrete GRF prior model given a sample data set was developed. These parameters describe the type of label image that will be favoured by the prior model. It is therefore important to choose a good set of GRF model parameters.

For the new parameter estimation method the idea of equivalencies was introduced in Section 8.2 based on the fact that the GRF model does not distinguish between some local configurations. These families of configurations that are not distinguished between are called equivalencies and can be treated as a single local configuration. Furthermore the total number of equivalencies is independent of the number of classes used in the model. An expression was derived giving the number of equivalencies and a method of remapping the discrete GRF model with regard to the equivalencies was defined. The equivalencies therefore reduce the GRF model to its true complexity and give an insight into the behaviour of the model. Note that equivalencies have been used before by Chen [1988] but were not investigated in any depth.

The new parameter estimation method is based on a cost function describing the difference between the local joint PDF of the model and its equivalent estimated from the sample data. The local joint PDF of the model however was impractical to estimate, due to the complexity of the partitioning function, and was therefore derived from the local conditional PDF of the model and the local joint PDF of the neighbourhood system as estimated from the data set. The resulting local joint PDF of the model and the data were both remapped and thereby greatly reduced in complexity. The squared difference between the PDFs was calculated and minimised with regard to the model parameters to give an estimate of the optimal parameter set.

In Chapter 10 this new parameter estimation method was tested and the results were compared to results from Besag's parameter estimation method (coding method in Besag [1974]). Two sets of experiments were performed in order to assess the quality of the estimated parameters.

In the first set of experiments, in Section 10.1, GRF realisations were created using the metropolis sampler. For the GRF realisations the model type, i.e. neighbourhood system and clique configuration, and the model parameters were known. The true model parameters could therefore be compared to the estimated ones.

Several different GRF realisations, of size 200×200 pixels, using a wide range of model types and parameters, were created. Then the new parameter estimation method and Besag's parameter estimation method were used to estimate the parameters of these samples. Comparing the parameter estimates from the two different methods to the true parameters, used in creating the GRF realisations, shows that both methods produce similar results. The new parameter estimation method performed slightly better than Besag's parameter estimation method. For both methods it is true that the higher the complexity of the GRF model (higher order neighbourhood and distinguishing between more clique types) and the larger the GRF model parameter values the less accurate are the parameter estimates. The reason for this is very likely the size of the GRF realisations. Larger sample sizes are required to represent more complex and more homogenous GRF model behaviour. The computational cost of generating and analysing larger GRF realisations was considered too great to justify this.

In the second set of experiments three test label images, two hand drawn with two classes and one extracted from a GIS data set with five classes, were used. The two hand drawn label images were of size 256×256 pixels and the label image created from the GIS data was 512×512 pixels. For all three test label images the GRF model parameters were estimated using different model types and clique configurations. For comparison Besag's method was also used.

It was clear that in some cases the parameter estimates were obviously too large (this can be concluded from experience with GRF models and their parameters). This was the case once for the new parameter estimation method and twice for Besag's parameter estimation method. It is not clear what the reason for this failure was.

It was impossible to directly assess the quality of these parameter estimates since the true GRF model parameters of the test label images were not known. However by using noisy versions of the test data and classifying this data, using the GRF model and the estimated parameters in a MAP

approach, the quality of the classification results could be assessed.

For the two hand drawn test label images single look intensity SAR data was simulated. For the five class label image from the GIS data set, true 9-look intensity AIRSAR data in the form of three bands (C, L and P) and each with three polarisations (HH, HV and VV) was available. Simulated annealing was then used to obtain the MAP estimates of the label images using all the different GRF model configurations and estimated parameter sets. The MAP estimates were then assessed by comparing them with the original test label images and giving the percentage of wrongly classified area. The assumption was that better GRF model parameters yield a better classification result.

For the first hand drawn test label image and its estimated parameters the number of wrongly classified area indicated, on average, better results for the parameters from Besag's parameter estimation method. For the second hand drawn test label image all results were better for the new parameter estimation method. For the third test label image, created from GIS data, the results indicated a slightly better fit for all but one of the parameter sets estimated using the new parameter estimation method.

For a further comparison the ML estimates (no prior model used) of the label images using the noisy test data were calculated. As expected these ML estimates gave results much worse than the worst MAP estimates. This indicated that using the prior model is advantageous no matter which of the two parameter estimation methods and what model configuration is used for these specific data sets.

In general it can be concluded that overall the new parameter estimation method sightly outperforms Besags's parameter estimation method and under certain circumstances one of the two parameter estimation methods may fail while the other still performs reasonably. If the test label image is small, a small neighbourhood and a single parameter for all cliques will yield better results rather than choosing a complex model configuration.

The computational cost of the new parameter estimation method is slightly higher than Besag's parameter estimation method but both can be performed in a very short time. The simplicity of Besag's parameter estimation method is a clear advantage. However the mapping of equivalences gives a deeper understanding of the underlying GRF model and can be useful in other applications.

11.3 SUGGESTIONS FOR FUTURE WORK

The MAP estimate based on the Bayesian approach was used in this thesis as a classification method. The two main parts of the MAP model are the likelihood function and the prior model and this thesis contributes to both areas. Firstly a new likelihood function describing three-polarisation multilook intensity SAR data was derived and surrounding issues such as parameter estimation were addressed. Then secondly the GRF was adopted as a prior model and a new parameter estimation method was developed enabling the parameter estimation from sample label images. A new understanding of the GRF model was possible due to the introduction of the equivalencies

which reduced the GRF model to its true complexity.

The two main contributions described can be seen as building blocks in the bigger picture of image analysis. The new likelihood function is very general in that it describes the averaged intensity samples from a three-dimensional complex zero-mean circular Gaussian process and should have other applications in areas unknown to the author. There are extensions to the models that could be explored. In SAR processing for example this model could be used as the speckle process part of a likelihood function, which should include additional models for ground cover texture and other statistical effects. In Section 7.3 the characteristic function was used to derive an alternative form of the joint density. This joint density contains an integral which was considered too difficult to solve. However numerical methods can be used to evaluate this PDF. Computational cost and accuracy between the two different joint densities could be analysed. Furthermore the development of approximations of the new likelihood functions in the form of skewed three-dimensional correlated Gaussian PDFs could be investigated in order to speed up the processing time. The cost function, developed in the parameter estimation method for GRF models, could be investigated with regard to generating GRF realisations by way of optimising the cost function with regard to the label image rather than the model parameters. Additionally other metrics in the cost function, such as the log-likelihood, could be tested and compared to the one given. The equivalencies could be used to generate much shorter lookup tables (independent of the number of classes) for the local PDF value of a label configuration in order to speed up the optimisation process during a MAP estimation.

The RADARSAT-2 mission, planned for launch in early 2004, in one imaging mode provides suitable data (averaged intensity values of the three polarisations rather than the full complex covariance matrix) [Staples and Hornsby 2002] for model analysis and development.

Above all it should be emphasised that the use of computers in image analysis with the continuously growing computing power gives the ability to develop more sophisticated image analysis systems. Much work and development is required in this area.

Appendix A

GEGENBAUER FUNCTION

The Gegenbauer function (also called ultraspherical polynom) belongs to the group of orthogonal polynomials and is described in Abramowitz and Stegun [1972] chapter 22. The Gegenbauer function in relation to the results given in this thesis is $C_{k-n+1}^{n-1}(t)$ which is the coefficient for the term $\mathcal{K}^{k-n+1}$ in the expansion of $(1-2\mathcal{K}t+\mathcal{K}^2)^{-n+1}$. This function can be evaluated in a finite sum

$$C_{k-n+1}^{n-1}(t) = \frac{1}{(n-2)!}\sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m (k-m-1)!}{m!(k-n+1-2m)!}(2t)^{k-n+1-2m} \tag{A.1}$$

where $\lfloor \cdot \rfloor$ is the floor operation rounding $\cdot$ to the next smaller integer. Note that $C_0^{n-1}(t) = 1$.

REFERENCES

M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 9th edition, 1972.

A. Baraldi and F. Parmiggiani. A refined gamma MAP SAR speckle filter with improved geometrical adaptivity. *IEEE Trans. Geosci. Remote Sens.*, 33(5):1245–1257, September 1995.

A. Barni, M. Betti, and A. Mecocci. Fuzzy segmentation of SAR images for oil spill recognition. *IEE Image Processing and its Applications*, (410):534–538, July 1995.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society*, 2:192–236, 1974.

J. Besag. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society B*, 48(3): 259–302, 1986.

C.-C. Chen. *Markov random fields in image analysis.* PhD thesis, Department of Computer Science, Michigan State University, 1988.

R. Cook, I. McConnell, and D. Stewart. Segmentation and simulated annealing. In *Proc. Europto Conf. on SAR Image Analysis and Modeling II, SPIE*, volume 2958, pages 30–37, Taormina, Italy, 1996.

J. C. Curlander and R. N. McDonough. *Synthetic Aperture Radar: System and Signal Processing.* Wiley Series in Remote Sensing. John Wiley & Sons, Inc., USA, 1991.

H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pat. Anal. Mach. Intell.*, PAMI-9(1):39–55, January 1987.

S. Derrode, G. Mercier, J. L. Caillec, and R. Garello. Estimation of sea-ice SAR clutter statistics from pearson's system of distributions. In *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01 Proceedings*, Sydney, Australia, July 2001.

R. C. Dubes and A. K. Jain. Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.

H. Espedal. Detection of oil spill and natural film in the marine environment by spaceborne SAR. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings*, volume 3, pages 1478–1480, Hamburg, Germany, 1999.

L. Ferro-Famil, E. Pottier, and J.-S. Lee. Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier. *IEEE Trans. Geosci. Remote Sens.*, 39(11):2332–2342, 2001.

R. Fjørtoft, F. Lebon, F. Sery, A. Lopes, P. Marthon, and E. Cubero-Castan. A region-based approach to the estimation of local statistics in adaptive speckle filters. In *Proceedings of IGARSS'96*, volume 1, pages 457–459, Lincoln, Nebraska, May 1996.

V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pat. Anal. Mach. Intell.*, PAMI-4(2):157–166, March 1982.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pat. Anal. Mach. Intell.*, PAMI-6(6):721–741, November 1984.

J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser speckle and related phenomena*, volume 9 of *Topics in applied physics*, chapter 2. Springer-Verlag, 1975.

J. W. Goodman. Some fundamental properties of speckle. *J. Opt. Soc. Am.*, 66(11):1145–1150, November 1976.

J. W. Goodman. A random walk through the field of speckle. *Optical Engineering*, 25(5):610–612, May 1986.

N. R. Goodman. Statistical analysis based on a certain multivariante complex Gaussian distribution (an introduction). *Ann. Mathemat. Statist.*, 34:152–177, 1963.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, 5th edition, 1994.

M. Hagedorn, P. Bones, D. Pairman, and H. North. Parameter estimation for prior markov random field models. In *Proceedings of Irish Machine Vision and Image Processing*, pages 253–256, Maynooth, Ireland, 2001.

M. Hagedorn, P. Smith, P. Bones, R. P. Millane, and D. Pairman. A trivariate chi-squared distribution derived from the complex Wishart. *Journal of Multivariate Analysis*, accepted for publication August, 2004.

Y. Hara, R. G. Atkins, R. T. Shin, J. A. Kong, S. H. Yueh, and R. Kwok. Application of neural networks for sea ice classification in polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 3(33):740–748, 1995.

D. W. Hawkins. *Synthetic Aperture Imaging Algorithms: with application to wide bandwidth sonar.* PhD thesis, Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand, 1996.

S. M. Kay. *Fundamentals of Statistical Signal Processing: estimation theory*, volume 1 of *Prentice-Hall Signal Processing Series*. Prentice-Hall, 1993.

S. Kotz, N. Balakrishnan, and N. L. Johnson. *Models and applications*, volume 1 of *Continuous multivariate distributions*. A Wiley-Interscience Publication, New York, 1972.

A. S. Krishnamoorthy and M. Parthasarathy. A multivariate Gamma-type distribution. *Ann. Math. Stat.*, 22:549–557, 1951.

S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. Pat. Anal. Mach. Intell.*, 11(8):799–813, August 1989.

J.-S. Lee. A simple speckle smoothing algorithm for synthetic aperture radar images. *IEEE Transactions on System, Man, and Cybernetics*, SMC-13(1):85–89, January-February 1983.

J.-S. Lee, M. Grunes, and R. Kwok. Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution. *Int. J. Remote Sensing*, 15(11):2299–2311, 1994a.

J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude. Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Trans. Geosci. Remote Sens.*, 37(5):2249–2258, September 1999.

J.-S. Lee, M. R. Grunes, and S. A. Mango. Speckle reduction in multipolarization, multifrequency SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 29(4):535–544, July 1991.

J.-S. Lee, M. R. Grunes, and E. Pottier. Quantitatve comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR. *IEEE Trans. Geosci. Remote Sens.*, 39 (11):2343–2351, November 2001.

J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller. Intensity and phase statistics of multilook polarimetric and inferometric SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 32(5):1017–1028, September 1994b.

J.-S. Lee, I. Jurkevich, P. Dewaele, P. Wambacq, and A. Oosterlinck. Speckle filtering of synthetic aperture radar images: a review. *Remote Sensing Reviews*, 8:313-340, 1994c.

J. A. Lewis, F. M. Henderson, and D. W. Holcomb. *Principles and Applications of Imaging Radar*, volume 2, chapter 3 Radar Fundamentals: The Geoscience Perspective. John Wiley & Sons, Inc, 3rd edition, 1998.

G. Liu, S. Huang, A. Torre, and F. Rubertone. Unsupervised multiresolution segmentation and interpretation of textured SAR images. In *Proceedings of the 3rd Conference on Image and signal processing for remote sensing III, SPIE*, volume 2955, pages 261–271, Bellingham, WA, 1996.

A. Lopès and F. Séry. Optimal speckle reduction for the product model in multilook polarimetric SAR imagery and the Wishart distribution. *IEEE Trans. Geosci. Remote Sens.*, 35(3):632–647, May 1997.

A. Lopes and R. Touzi. Adaptive speckle filtering for SAR images. In *Proceedings of IGARSS'88*, pages 1263–1266, Edinburgh, Scotland, September 1988.

K. S. Miller, R. I. Bernstein, and L. E. Blumenson. Generalized Rayleigh processes. *Quart. of Appl. Math.*, 16:137–145, 1958.

H. Mott. *Antennas for Radar and Communications: A Polarimetric Approach.* Wiley-Interscience, 1992. ISBN 0471575380.

J. D. C. Munson and J. L. C. Sanz. Image reconstruction from frequency-offset Fourier data. *Proceedings of the IEEE*, 72(6):661–669, 1984.

E. Nezry, A. Lopes, and R. Touzi. Detection of structural and textural features for SAR images filtering. In *Proceedings of IGARSS'91*, volume 4, pages 2169–2172, Espoo, Finland, June 1991.

J. A. Ogilvy. *Theory of wave scattering from random rough surfaces.* IOP Publishing Ltd, 1991. ISBN 0-7503-0063-9.

A. Papoulis. *Probability, random variables, and stochastic processes.* McGraw-Hill, 3rd edition, 1991.

R. K. Raney. *Principles and Applications of Imaging Radar*, volume 2, chapter 2, *Radar Fundamentals: Technical Perspective.* John Wiley & Sons, Inc, 3rd edition, 1998.

E. Rignot and R. Chellapa. Segmentation of polarimetric synthetic aperture radar data. *IEEE Trans. Image Process.*, 1(3):281–300, July 1992.

S. Saatchi and M. Moghaddam. Estimation of boreal forest biomass using spaceborne SAR systems. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings*, volume 3, pages 1646–1648, Hamburg, Germany, 1999.

G. C. Staples and J. Hornsby. Turning the scientifically possible into the operationally practical: RADARSAT-2 polarimetry applications. In *Proceedings of IGARSS 2002*, volume 3, pages 1474–1476, Toronto, Canada, June 2002.

H. Stark and J. W. Woods. *Probability, Random Processes, and Estimation Theory for Engineers.* Prentice-Hall, Inc., 2nd edition, 1994. ISBN 0-13-728791-7.

H. Stark and J. W. Woods. *Probability and Random Processes with Application to Signal Processing.* Prentice-Hall, Inc., 3rd edition, 2002. ISBN 0-13-020071-9.

D. Thompson. The potential of SAR interferometry for oceanographic measurements: a review. In *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01 Proceedings*, volume 1, pages 573–574, Sydney, Australia, July 2001.

F. T. Ulaby. *From theory to application*, volume 3 of *Microwave remote sensing*. Addison-Wesley Publishing Company, 1982a.

F. T. Ulaby. *Radar remote sensing and surface scattering and emission theory*, volume 2 of *Microwave remote sensing*. Addison-Wesley Publishing Company, 1982b.

F. T. Ulaby and C. Elachi. *Radar polarimetry for geoscience applications*. Artech House, Inc., 1990. ISBN 0-89006-406-7.

G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction*. Springer, 1995.

INDEX